SMILE AND TALK - ANA, THE ARCHITECTURE OF A ROBOT THAT VERBALLY AND NONVERBALLY UNDERSTANDS YOU

PAULO G PINHEITO[1], JOSUÉ J G RAMOS[2], GUSTAVO H DE OLIVEIRA[2]

[1]*Hoobox* , [2]*NRVC(Centro de Tecnologia da Informação Renato Archer)*
Rod D Pedro I Km 143,6 13069-901, Campinas, SP, Brazil
paulo@hoobox.com, josue.ramos@cti.gov.br, gustavo.oliveira@cti.gov.br

Abstract— Robots capable of recognizing and analyzing verbal and nonverbal communication can play important role in many domains, especially where the verbal communication does not provide enough information to build the best interaction experience. The ability to remember a person or to analyze his/her facial expressions can be important tools to improve the human-robot interaction, especially when the robot might have to deal with different people for long time. In this work it is presented the interaction architecture of ANA, the robotic receptionist created to interact with Brazilians speakers. The architecture provides support to verbal communication, such as voice and text inputs, and nonverbal information, such as facial and basic emotions recognition. To increase the robustness of interactions, the knowledge of the robot was modeled over ontologies using local and remote bases. To validate the model, ANA interacted with workers, provided information and captured their emotions and classified them. These sentiments were used to calculate the emotion average rate for each hour of the day, providing data about the workers' behaviors.

Keywords— Interaction, Verbal and Non verbal Communication, Emotion-Oriented System, Facial Expressions, Robotics

Resumo— Robôs capazes de reconhecer e analisar a comunicação verbal e não verbal podem desempenhar um papel importante em muitos domínios, especialmente onde a comunicação verbal não fornece informações suficientes para criar a melhor experiência de interação. A capacidade de se lembrar de uma pessoa ou de analisar suas expressões faciais podem ser ferramentas importantes para melhorar a interação humano-robô, especialmente quando o robô pode ter que lidar com pessoas diferentes por muito tempo. Neste trabalho é apresentada a arquitetura de interação da ANA, a recepcionista robótica criado para interagir com brasileiros. A arquitetura fornece suporte à comunicação verbal, como entradas de voz e texto, e informações não-verbais, como reconhecimento de emoções faciais básicas. Para aumentar a robustez das interações, o conhecimento do robô foi modelado em ontologias usando bases locais e remotas. Para validar o modelo, ANA interagiu com trabalhadores, forneceu informações e capturou suas emoções e as classificou. Esses sentimentos foram usados para calcular a taxa média de emoção para cada hora do dia, fornecendo dados sobre os comportamentos dos trabalhadores.

Palavras-chave— Interação, Comunicação verbal e não verbal, Sistemas orientados à emoção, Expressões Faciais, Robótica

1 Introduction

This work is motivated by the desire of having a robotic receptionist able to understand a person through verbal and nonverbal information (Ramos et al., 2015). In the receptionist services domain, the human-human interaction involves not only the exchange of verbal information, but also includes nonverbal information, such as the receptionist's memories about who that person is, or how often that person went through, or his/her most common question (Sokanu, 2017). Even information about the emotional state of the person can be used by the receptionist to provide the best experience on the current and next interactions. We believe this scenario will be the best for human-robot interaction as well.

In this domain, the platform for data capture and data persistence, besides communication strategies to support an efficient service, is very important. Hence, it is pursued to build a solid framework that could integrate all these parts. Nonverbal communication plays an important role to attract people and to maintain a human-robot interaction over long periods of time. Also, nonverbal communication may help on the fine-tuning of the interaction, for instance, not changing the information answered by the robot but changing the intonation on answering according to the mood of the person. On the other hand, verbal communication is directly related to efficiency of the service. Once the robot receives a verbal request for information, it must be able to understand and promptly answer to it.

In this work it is presented an architecture for human-robot verbal and nonverbal interactions that supports data persistence and access to local and global knowledge bases. The architecture counts on two rules machines, one for each type of interaction, which together create the responses to interactions. The robot consists of an avatar able to speak and to respond non verbally to requests as well. The motivation for this research is to implement this architecture in a robotic receptionist to be employed in Brazil (Ramos et al., 2015).

This paper reports our results from an experiment designed to explore the practical use of the receptionist robot ANA at the Division of Robotic and Visual Computing, in CTI, interacting with the workers during the fifteen days. Figure 1 shows the default setup of ANA.

The experiment has as main purpose to demonstrate the practical use of the developed platform on supporting verbal communication, recognizing speech and texts, being able to express itself using a synthetic speech, and supporting nonverbal communication, recognizing the user and her/his emotions, quantifying them and to provide data for potential analysis

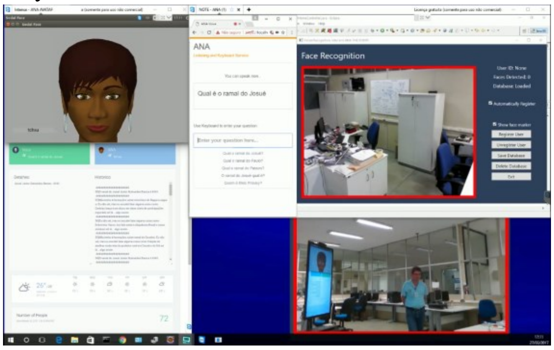

Figure 1. ANA, the receptionist robot at the Division of Robotic and Visual Computing in CTI (Centro de Tecnologia da Informação Renato Archer) in Campinas, Brazil

This paper is divided on the following sections: Section 2 Presents Verbal and Nonverbal Communication Systems requirements, Section 3 presents the Experimental Platform detailing its architecture and software components, Section 4 presents one experiments performed with platform and Section 5 presents the conclusions.

2 Verbal and Nonverbal Communication Systems

Many applications require verbal and nonverbal interaction capabilities. Assistive robots as companion for special group of people, such as elderly (Wada and Shibata, 2007) or children with autism (Kozima et al., 2005), robotic educational assistants (Robins et al., 2005), and robotic recepcionist, like the Arabic (Makatchev et al., 2010) and our Brazilian one (Trovato et al., 2015b) make part of this group.

According to Mavridis(2015), all these systems share the desirability of natural fluid interaction with humans supporting natural language and non-verbal communication. In this work is presented an overview of human-robot interactive communication, covering verbal and nonverbal aspects of human-robot interaction, bringing up a system might that have to meet some these requirements. Even though the list is neither totally exhaustive nor orthogonal, it is considered a good starting point.

For reference, the list of key points is composed of: breaking the “simple commands only" barrier - situations where only the human drives the conversation with a single-initiative dialogue, multiple speech acts and mixed initiative dialogue, motor correlates and nonverbal communication, purposeful speech and planning, multilevel learning, utilization of online resources and services, and finally, miscellaneous abilities. It can be seen that the proposed architecture in this paper attempts to objectively surround most of these points.

Affective markers, such as facial expressions or body postures, are also very important aspects in human-human interactions (Picard, 2003), and therefore should be considered in human-robot interaction as well. Cynthia Breazeal presented the Kismet (Breazeal, 1999), the robot capable to perform multiple facial expressions while interacting with people. Some other Works focused on emotion recognition, such as Kaliouby and Robinson (2004); Bartlett et al. (2005) to improve the interaction.

Text-to-speech and speech recognition are also key points, not only for verbal communication but to improve the affective link (Schrder, 2009). In our previous work (Trovato et al., 2015b) different types of voices for the robot were tried out in interactions with Brazilian public. In this study, it was observed that a smooth voice was preferred. This voice is the one will be used in the proposed model.

In general, to provide a complex service as to require verbal and nonverbal interactions, a knowledge base and the ability to represent and manipulate knowledge is mandatory (Saxena et al., 2014). Nowadays, it can be seen many robotic applications combining information at a large-scale and utilizing online information in order to enhance its communication abilities and to provide a better service. For instance, Facebots, the physical robots that utilize and share information on Facebook, creating dialogues towards enhancing long-term human-robot interaction (Mavridis et al., 2010). This work describes the example of using knowledge bases, where the robot is capable to collect and to represent a large amount of information using ontology over local and remote sources.

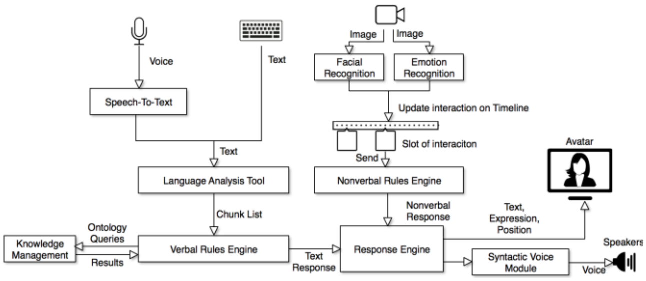

Figure 2: Architecture Overview.

3 Experimental Platform

Our research platform is ANA, a robotic receptionist designed for social interaction supporting verbal and nonverbal communication (Trovato et al., 2015b). ANA plays the role of receptionist of the department welcoming passersby and giving them information about general questions, number extension, schedule of the department as well as delivering personal messages. This robot version is built upon the receptionist robot originally developed by Prof. Reid Simmons's team at Carnegie Mellon University (Kirby et al., 2005). The first figure of this paper (Figure 1) shows the avatar of ANA integrated to the proposed model and Figure 2 the architecture.

Besides the receptionist services, ANA presents daily news and the weather forecast. Such information attract the attention of the workers increasing the number of interactions. These attractions are care-

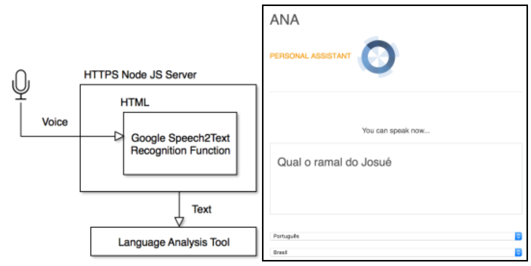

Figure 3:

a) Google Speech-To-Text Recognition function constantly running on a local and secure server..

(b) HTML page showing the text in Portuguese of the processed speech: ``What is the phone extension of Josué"

fully chosen to be neutral and not to influence the employee's humor. The experimental platform consists of the avatar of ANA, the proposed model and sensors which will be mentioned in this work.

3.1 Verbal Communication

3.1.1 Voice Input

It is used Google Speech API embedded on HTTPS server to provide an URL address and to maintain the permission to access the microphone without asking for it on future accesses. The usage of HTTP Secure protocol ensures the service will be standing for long periods of time. Figure 3a shows the speech to text module in more details. Figure 3b shows the web page that displays what the user is speaking in real-time. Although it has been shown to be the best solution so far, some failures can be found, such as the fact the service is not able to deal with interrogation and other marks.

Once the speech input is completed by the user, the recognizer processes it and generates the corresponding sentence. The recognizer understands when the user stopped talking due to a prolonged speechless phase. The corresponding sentence is sent to the Language Analysis Tool that is a thread permanently ready to receive text sentences, as it will be shown in Section 3.1.3

3.1.2 Text Input

Text inputs are provided by the keyboard placed near to the robot. The user can type anything and send it by pressing enter. In this case, the inputs are sent directly to the Language Analysis Tool.

3.1.3 Language Analysis Tool

Before applying the main rules of interaction, the language analysis tool makes the lexical treatment of the verbal input breaking the sentence into lexical chunks and then creating the lexical map. As an example it can be used the sentence recognized on the Speech to Text process shown, in Figure 3b , ``Qual o ramal do Josue" that means in Portuguese ``What is the extension number of Josue". The lexical map given by the analysis tool is:

Sentence: ``Qual o ramal do Josue".

Lexical Map: qual(P0CS0000), o(DA0MS01), ramal(NCMS0002), de(SPS003), o(DA0MS04), josue(NP000005).

Six chunk are returned: i)`qual' is a Pronoun (P), used interrogatively as a request for specific information; ii) `o' is the Definite Article (DA); iii) `ramal' is Singular Masculine Common Noun (NCMS); `do' - is broken into two chunks, since in Portuguese, `do' it is a contraction of the iv) joint preposition `de' and the v) definitive article `o'; and finally, vi) `josue', that is a Proper Noun (NP).

3.2 Verbal Rules Engine

Given the lexical map, this engine is responsible for generating the verbal response for the user verbal request. For the receptionist domain, this module consists of four basic rules that are triggered:

-Rule 1: If there is at least one Proper Noun (PN) and one Common Noun (CN).

-Rule 2: If there is only Common Nouns (CN) or only Proper Nouns (PN).

-Rule 3: If passed through the Rule 2, but the system created no sufficient answer.

-Rule 4: If the case was neither covered by Rule 1, 2 nor 3.

3.2.1 **Rule 1: PN and CN**

This rule is triggered when there is a sentence like this:

Request format: “What is the [Common Noun] of [Proper Name]?".

Request example: “What is [the extension number] of [Josue]?"

Proper Nouns are related to ontology's individual resources, and Common Nouns to property's resources. If there is a proper name, the engine searches over the ontology for the individual who has that proper name value. Finding the ontology's individual, the engine looks for his/her property that is given by the common noun. Then is created the answer:

Response format: “The [Common Noun] of [Proper Name] is {Value(Proper Name, Common Noun}”.

Response example: ``The [extension number] of [Josue] is 3882}".

3.2.2 **Rule 2: Only PN or Only CN**

This rule is triggered when there is a sentence like:

Request format: “Who is the [Proper Name]”'.

Request example: ``Who is [Josue]?".

Or

Request format: “What is [Common Noun]?”.

Request example: “What is a [Robot]?”.

When there is only a Proper Name (PN), the engine searches over the ontology for individuals pushing its super-classes.

Response format: “[Proper Name] is a [Super-Classes]”

Response example: “[Josue] is [a Supervisor, Teacher and Employee]”.

If the super-class was the root node (Thing), the description of individual is used instead:

Response format “[Proper Name] is [description]”.

Response example: “[Josue] is [the one who work at the Robotic Division and plays basketball.]”

When there is only Common Noun, it looks for properties. Then the module gets the individuals who haverelation with, then its super-class:

Response format “[Common Noun] is a property of [Individual]”.

Response example: “[Extension number] is a property of [Employee]”.

In case the super-class is the root node, the description of the property will be considered instead:

Response format “[Common Noun] is [description]”.

Response example: “[Extension number] is [the number you call when you want to talk to somebody.]”

3.2.3 **Rule 3: From the rule 2, the created answer is insufficient**

Passing through the Rule 2, the module might generate an answer that is not satisfactory. It is defined as satisfactory one, the answer that returns at least one of the super-class of the individual (but not the root) or her/his description. For property, a satisfactory response is the one that returns its description or at least a super-class of the individual who owns that attribute

Insufficient Responses Examples:

“[Josue] is a [Person].”

“[Josue] is [no description available].”

“[Extension number] is a property of [Person].”

“[Extension number] is [no description available].

When the response is insufficient for individuals, the engine list some other individuals who are part of the same class. When the response is insufficient for an attribute, the external DBPedia (Morsey et al., 2011) ontology is consulted.

Bypassing Insufficient Responses:

“[Josue] is [a Person] like [Samuel, Peter and John].”

“[Extension number] is [an additional telephone wired to the same telephone line as another.

3.2.4 **Rule 4: There is no sufficient data for a Response**

Even after consulting the local and DBPedia ontologies, if there is no sufficient answer for the request, a generic response is given:

Generic Response:

“ I do not know anything about it.”

3.3 Knowledge Management

Ontology is a term that refers to the shared understanding of some domain, that can be used to unify different frameworks to solve common problems (Uschold and Gruninger, 1996). An ontology usually encompasses the knowledge about the world, representing it through concepts of class, entities, attributes and processes trying to eliminate conceptual and terminological confusion about the world. Once the knowledge is modeled, relationships can be set out. In fact, these networks of relationships, the consistency and lack of ambiguity are important characteristics of a modeling ontology system.

In general, to use an ontology as knowledge representation, some steps must be followed. The first one is identifying the purpose and scope. Here, it is defined that the receptionist robot should be aware of common terms of a receptionist and of people and projects that are running in the department.

The second one is building the ontology. When ontologies get bigger, the use of management tools is required for ontology verification and validation, ensuring that particular constraints will be respected (Jurisica et al., 2004). For this project it is used the Protégé tool (Horridge et al., 2006). Protégé is an open source OWL ontology editor that uses the OWL API for management tasks, from loading to saving them.

The third one is coding and integrating existing ontologies into the system. It is represented by accessing and processing the ontology. In this work we have used Jena2 ontology API (Carroll et al., 2004). The ontology was created on Protégé editor and the system accesses it using this API. Two ontology basis will be used. A local one, consisting of information about our specific domain, and the online ontology, DBPedia (Morsey et al., 2011).

3.3.1 How it works

The Verbal Rules Engine, as it is mentioned in the Section 3.2, is the module that access the ontology, focusing on classes, individuals and attributes. Into the engine, a module is created pointing out to the OWL file (the representation of the ontology). When a query is requested (e.g. to figure out if ‘Josue’ is an individual, a class or attribute), a resource with the name `Josue' is created and then, using the API, it is possible to discover what the classification is.

Back to example of the Rule 1 (Section 3.2), the system knew by the lexical map that `Josue' is a Proper Name, helping the engine to search for `Josue' directly on individuals. The search is made by asking to the ontology if the resource created 'Josue' is an Individual or not. Similar process is done with attributes. The lexical map helps the system to look for `the extension number' directly on the properties, saving time avoiding the search over other classifications. Alike, the research is done by creating the resource `the extension number' and asking the ontology whether that resource exists or not, its classification and the value of the relationship, as follows:

Request:

“What is [the extension number] of [Josue]?”

Ontology Confirmation:

property: ramal individual: Josue

Ontology Return:
property:ramal [of] individual:Josue [is] {3882}.

3.4 Nonverbal Communication

In this work, the nonverbal communication is restricted to user presence, facial and emotion recognition. ANA is able to use the information provided from these recognition inputs to improve its answers. For instance, using the person's name gathered from facial recognition to call him/her by the name, or using the emotion captured to talk to the user in a more smooth or harsh way.

To capture the emotions and to perform facial recognition, it is used Intel RealSense™ 3D Camera (F200) (Intel, 2017) This is a stand-alone camera that uses depth-sensing technology and can be attached to a desktop or laptop computer, extracting emotions and performing facial recognition in real time. It consists of a conventional camera, an infrared laser projector, an infrared camera, and a microphone array. The camera features facial analysis tracking 78 points on the face inferring emotions and sentiments. To our knowledge, ANA is the first project to use this technology as a tool for acquisition of emotions to be used in a Human-robot interaction system. Figure 4 shows the camera and the capture of emotion and facial recognition.

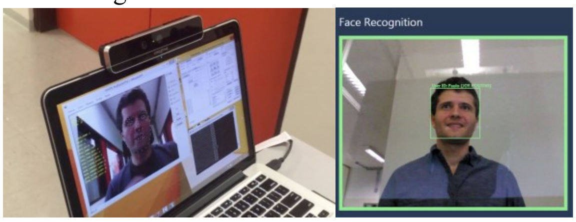

Figure 4:
a) Camera Intel F200 - (b) Face and EmotionRecognition.

The emotion recognition algorithm supplied in Real Sense SDK supports 6 primary emotions: Anger, Disgust, Fear, Joy, Sadness, Surprise. It is applied a filter to extract the basic sentiments, categorizing them into negative, positive or neutral sentiment. On the first line of Figure 5a , there is neutral expressions, and on the other lines, expressions that are recognized as positive sentiment.

On the first line of Figure 5b, there is another set of neutral expressions and on the other lines, expressions recognized as negative sentiment.

The system categorizes the emotions of an interaction into these three major groups. Let's suppose the user has interacted for 20 seconds with the robot that captured 10 expressions: {NEUTRAL, NEUTRAL, POSITIVE, POSITIVE, NEUTRAL, NEUTRAL, NEGATIVE, NEGATIVE, NEGATIVE, NEGATIVE}. The balance of this interaction is -2. The reverse applies to positive balances and a neutral humor interaction means the balance of negative and positive emotions is zero. There is no minimum time to define an interaction. An interaction will last the time the user spends in front of

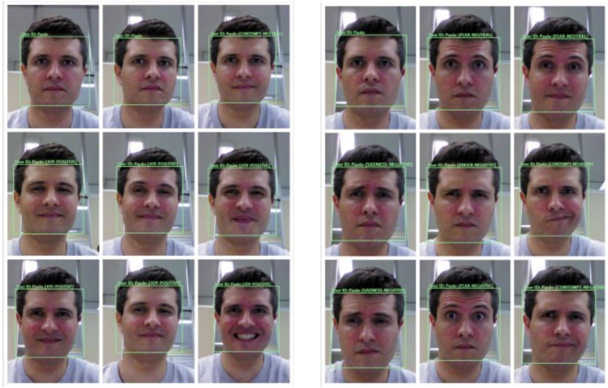

Figure 5:
(a) Neutral faces on the first line and examples of positive sentiment
(b) Neutral faces on the first line and example of negative sentiment

the robot. The final emotion balance is saved into the interaction slot.

3.5 Timeline - Data From Human–Machine Interaction

Data collection has been a challenge when involves nonverbal parts such as emotion in conversation (Bickmore and Picard, 2005). For many applications, video recordings, transcription and analysis are necessary and are time consuming. In this work we do not use inputs that must be treated offline such as video recordings. Rather, the capture of emotions and facial recognition data are performed in real-time. To make this possible, it is used the concept of Timeline to organize a list of events in chronological order. Figure 6 shows the representation of Timeline.

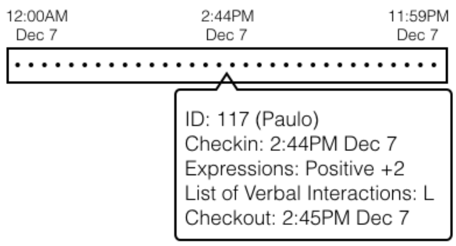

Figure 6: Timeline representation and a data slot of an interaction.

Each interaction creates an event (slot). Each slot is saved on Timeline at the point corresponding to the time it happened. Emotions captured from the user and its recognition data are inserted into that slot as well as all verbal information. When the user goes away, the slot is close and saved. To retrieve this information, the system loops back through the Timeline structure pushing the data.

3.6 Nonverbal Rules Engine

In Section 3.2, it is described how the Verbal Rules Engine deals with verbal requests. Nonverbal Rules Engine works similarly but being triggered by nonverbal stimuli. In this work it is considered: user presence, facial and emotion recognition. This engine is made up of three rules groups:

-Checking-in Rules - Rules triggered when the user appears on the robot's field of view.

-Checking-out Rules - Rules triggered when the user

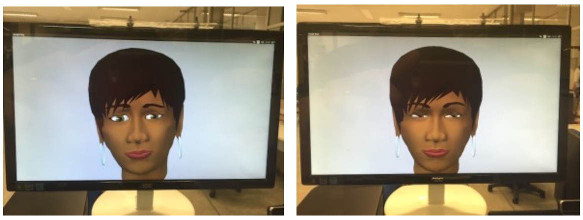

Figure 7:
a) ANA performs a sad face when the user leaves.
(b) ANA `sleeps' while no one appears.

steps out from the robot's field of view.
-On-Interaction Rules - Rules triggered while the user is interacting.

3.6.1 Checking-in Rules

As described in the Section 3.5, when the user appears on the robot's field of vision, the slot in the Timeline is created for that interaction. Based on the facial recognition, the system can tell how many times the user checked in. If it is the user's first time, a first time welcome message is triggered, if it is not, a general welcome message is the chosen one:

Welcome Message for First Time Checking in:
``Josue, nice to see you here for the first time."\\

Welcome Message for Non-First Time Checking in: ``Josue, nice to see you again."

In case the user is not recognized by the system, the same message is triggered, but without the name reference. In addition to the spoken welcome message, a subtitle of it pops up. The system can behave using the same emotion expression the user is performing. If the person features a positive emotion, the system can speak the welcome message performing a positive expression of joy.

3.6.2 Checking-out Rules

When the user steps away from the robot's field of vision, a facial expression of sadness is triggered. Non verbal communication (spoken message) is performed by the robot. Figure 7 shows the avatar behavior when the user goes away.

3.6.2 On-Interaction Rules

During the interaction, the robot tries to keep eye contact tracking the user head position. The second rule is triggered when the user remains for a long time in front of the robot without interacting. In this case, ANA verbally asks if the user needs any help. The third rule consists on the robot trying to mimic the basic user sentiments (positive, negative and neutral). Figure 8 shows the avatar behavior while the user is moving around.

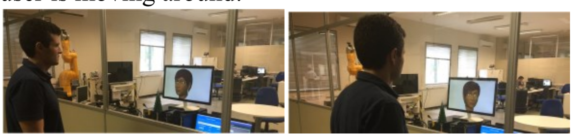

Figure 8: ANA starring the user while he moves from the left to the right.

3.7 Response Engine

The robot's response to the user's request may be the union of several simultaneous responses created by verbal and nonverbal engines. Figure 9 illustrates the combination of responses managed by the Response Engine.

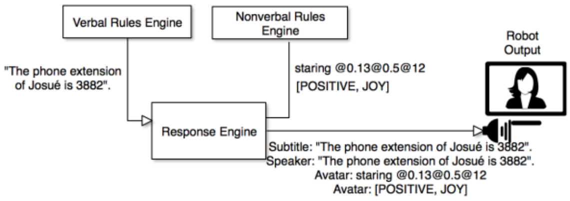

Figure 9: Response Engine managing verbal and nonverbal responses for a request and forwarding them on appropriate format to the output devices.

In this example, the user asked to the robot ``What is the phone extension of Josue". The Verbal Rules Engine brought out the textual answer and the Nonverbal Rules Engine captured the position of the user head and emotion. Then the information is formatted and channeled to the appropriate output devices.

3.8 Robot Output

The robot interaction interface consists of the avatar, subtitles and audio output. The textual response sent by the Response Engine are forwarded to the voice synthesizer, that plays the audio through speakers. The engine sends the text to the avatar, that moves the lips in accordance while displaying it as subtitles. The avatar also receives nonverbal information about the user emotion and the head position. Figure 10 shows the distribution of these message.

3.8.1 Avatar

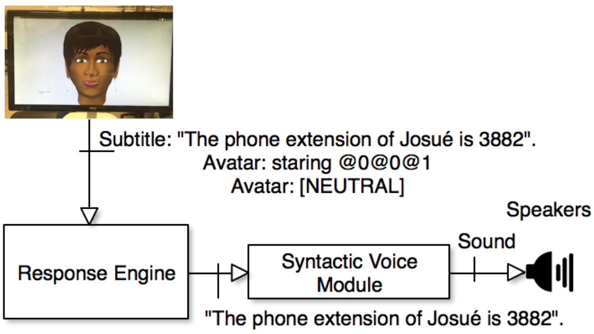

Figure 10: System output, the avatar's display and the syntactic voice module

ANA's avatar is an inherited system from CMU (Carnegie Mellon University) originally used as a receptionist robot at University (Kirby et al., 2005). The avatar is capable of performing dozens of facial expressions, moving lips according visemes words, and capable of displaying a small caption on

Figure 11: Some of the expressions that the avatar can perform

the screen. Figure 11 shows some of the expressions most frequently used by ANA in our system. In Brazil, the robot had its physical characteristics changed to suit the country's culture according to research conducted by our group on preferences and cultural patterns (Trovato et al.,2015b) .

In order to design the receptionist robot to be employed in Brazil, the authors in (Trovato et al., 2015a) conducted a research about what the expectation of people is regarding the voice of the robot. Even though the accuracy of information is more important than the voice, it should be soft. The Syntatic Voice Module uses a neutral pitch with female voice to deliver the soft effect pointed out by the mentioned work. The speech is output by speakers placed close to the robot.

4 Experiments

The experiment was designed to test and demonstrate the practical use and the robustness of the developed platform on:

-Supporting verbal communication, recognizing speech and text inputs, being able to express itself using a synthetic voice.

-Supporting nonverbal communication, recognizing the user and her/his emotions, quantifying them and to provide data for potential analysis.

For the experiment, the robot was setup at the main entrance of the Division of Robotic and Visual Computing in CTI (Centro de Tecnologia da Informação Renato Archer) in Campinas, Brazil. The robot interacted with the workers from 10 AM to 5 PM during fifteen days of October. Even if the worker did not verbally interact with the robot, once his/her face was detected, the expressions were captured and an interaction counted. The system was set out to be capable to provide information about extension numbers, people and projects, following the rules described on Section 3.2 and Section 3.6.

The system categorizes the emotions into the three major groups mentioned before: neutral, positive and negative sentiment. Since it is not looked for a fine grained emotion average and given the system is used for long term (days, weeks), these categories might be enough to extract the information necessary for this specific experiment.

4.1 Results

During 15 days, 138 verbal interactions took place. The interactions took 202.6 minutes in total. During the verbal interaction, nonverbal information was also captured. For each 1 second of interaction, the system can capture 5 facial expressions, on average. That means 60,780 facial expressions captured.

Figure 12a shows the amount of verbal interactions for each day of the experiment, and Figure 12b, the sum of all verbal interactions' time per day.

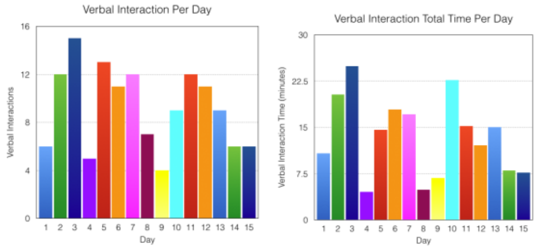

Figure. 12:

(a) Number of verbal interactions per day.

(b) Sum of the time (in minutes) of all verbal interactions per day

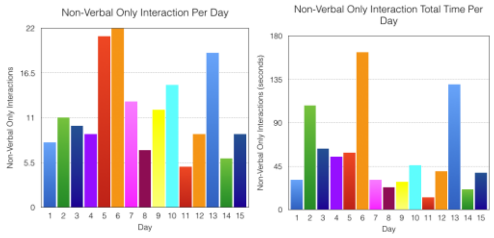

Figure 13:

(a) Number of verbal interactions per day.

(b) Sum of the time (in minutes) of all verbal interactions per day

Figure 13a shows the amount of nonverbal-only interactions for each day, and Figure 13b the sum of the time of all nonverbal-only interaction per day. The nonverbal-only interaction is characterized when the user does not speak to the robot, but appears in its field of view facing it.

During 15 days, 176 nonverbal-only interactions took place.

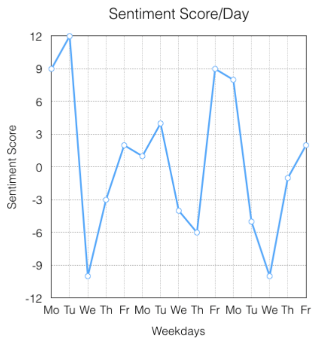

Figure 14: Sentiments.

The sum of time of all interactions was 850 seconds. Approximately 4,250 facial expressions were captured. Summing all the captured facial expressions (on verbal and nonverbal interactions) the system captured around 65,000 facial expressions. Figure 14 shows the 15 weekdays and the sentiment score average per day.

The experiments showed that the system appears to be robust for daily use, being able to interact with users in a verbally and non-verbally way. The time of nonverbal-only interactions (interactions where the user does not talk to the robot) were shorter, taking on average, 4.82 seconds, than verbal interactions, taking on average, 1.46 minute.

Given the results and the period of time, we cannot state any relationship between the number of verbal interactions and nonverbal interactions. We also cannot state a relationship between the duration time or number of interactions in relationship to the days of the week. However, on monitoring the workers' emotions, the results might point out weekly pattern. Considering the sentiment average score:

-Mondays and Fridays were always positive days.

-Wednesdays and Thursday were always negative days.

-If we consider an intermediate range between the value -3 and +3, we would have exactly 5 days with score above the range, 5 days under the range, and 5 days within the range.

Project approvals, important meetings, weather or even the restaurant day menu are some events that may be able to change the sentiment of the people, but this work did not consider these aspects.

Verbal rules applied here seemed to be a good fit for the receptionist domain where the questions asked by users were answered without causing apparent frustration. The nonverbal rules fulfilled the mission of attracting users to interact with the robot for longer duration.

5 Conclusions

In this paper It was presented the interaction architecture of ANA, the robotic receptionist, and its validation use for monitoring emotions during work days. The architecture provides support to verbal communication, such as voice and text inputs, and nonverbal information, such as facial and basic emotions recognition.

Acknowledgments

This Project is partially Sponsored by Fapesp Project 2013/26453-1 and CNPq-CTI/PCI 454796/2015-3.

References

Bartlett, M. et al. (2006) ‘Fully automatic facial action recognition in spontaneous behavior’. *In: Automatic Face and Gesture Recognition*, 2006. FGR 2006. 7th International Conference on. pp. 223-230.

Bickmore, T., Picard, R. (2005). ‘Establishing and maintaining long-term humancomputer relationships’. In: *ACM Trans Comput Hum* Interact (ToCHI). Vol. 59. pp. 21-30.

Breazeal, C. (1999).’Robot in society: Friend or appliance’. *In: Proc. Autonomous Agents Workshop on Emotion-Based Agent Architectures*. pp. 18-26.

Carroll, J. J., Dickinson, I., Dollin, C., Reynolds, D., Seaborne, A., Wilkinson, K. (2004). ‘Jena: Implementing the semantic web recommendations’. *In: Proceedings of the 13th International World Wide Web Conference on Alternate Track Papers &Amp*; Posters. WWW Alt. '04. ACM, New York, NY, USA, pp. 74-83..

Horridge, M., Tsarkov, D., Redmond, T. (2006) ‘Supporting early adoption of owl 1.1 with protege-owl and fact’,. *In: Proceedings of the Second OWL Experiences and Directions Workshop* (OWLED.

Intel, R. (2017). ‘Intel realsense 3d camera (f200)’. In:Developing for the Intel RealSense Camera URL https://software.intel.com/en-us/realsense/f200came

Jurisica, I., Mylopoulos, J., Yu, E. (2004). ‘Ontologies for knowledge management: An information systems perspective’, *Knowledge and Information Systems* 6 (4), pp. 380-401.

Kaliouby, R. E., Robinson, P.(2004).,’Real-time inference of complex mental states from facial expressions and head gestures’. 2012 *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops* 10, 154.

Kirby, R et al (2005). ‘Design-ing robots for long-term social interaction’. In*: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems* (IROS 2005). pp.2199-2204.

Kozima, H., Nakagawa, C., Yasuda, Y. (2005),’ Interactive robots for communication-care: A case-study in autism therapy’. *Robot and Human Interactive Communication - ROMAN* 1,pp. 341-346.

Makatchev, M et al .(2010). ‘Dialogue patterns of an arabic robot receptionist’. In: 2010 5th ACM/IEEE International Conference on *Human-Robot Interaction (HRI),*. pp. 167-168.

Mavridis, N., (2015), ‘A review of verbal and non-verbalhuman-robot interactive communication’. *Robotics and Autonomous Systems* 63, pp. 22-35.

Mavridis, N., Petychakis, M., Tsamakos, A., Toulis, P.,Emami, S., Kazmi, W., Datta, C., BenAbdelkader, C., Tanoto, A. (2010). ‘Facebots: Steps towards enhanced long-term human-robot interaction by utilizing and publishing online social information’. *Paladyn1* (3), 169-178.

Morsey, M., Lehmann, J., Auer, S., Ngomo, A. N. (2011).‘Dbpedia SPARQL benchmark - performance assess-ment with real queries on real data’. In: *The Semantic Web - ISWC 2011 - 10th International Semantic Web Conference, Bonn, Germany*, October 23-27,2011, Proceedings, Part I. pp. 454-469.

Picard, R. W. (2003). ‘Affctive computing: Challenges’. *International Journal of HumanComputer* Studies 59.

Ramos, J. G., Azevedo, H., Moroni, A., Trovato, G.,, Magossi, S. (2015). ‘Informações não-verbais na interaohumano-rob aplicado a um rob recepcionista’. In:*Simposio Brasileiro de Automacao Inteligente (SBAI*). p. available on researchgate.net.

Robins, B., Dautenhahn, K., Boekhorst, R., Billard, A (2005). ‘Robotic assistants in therapy and education of children with autism: can a small humanoid robot help encourage social interaction skills?’, *Universal Access in the Information Society 4 (2),* pp. 105-120.

Saxena, A., Jain, A., Sener, O., Jami, A., Misra, D. K.,Koppula, H. S. (2014). ‘Robobrain: Large-scale knowledge engine for robots’. In arXiv:1412.0691[cs.Ai].

Schrder, M. (2009). ‘Expressive speech synthesis: Past,present, and possible futures’. *In: Tao, J., Tan, T.(Eds.), A ective Information Processing. Springer London*, pp. 111-126.

Sokanu, C., (2017). ‘What does a receptionist do? In:What does a Receptionist do?’, URL https://www.sokanu.com/careers/receptionist/

Trovato, G. et al (2015a).’Designing a receptionist robot: Effect of voice and appearance on anthropomorphism’. *In: ROMAN. IEEE*, pp. 235–240.

Trovato, G. et al (2015b).’ Ola, my name is ana: a study on brazilians interacting with a receptionist robot’. *In: 17th International Conference on Advanced Robotics, ICAR*, pp. 66-71

Uschold, M., Gruninger, M. (1996). ‘Ontologies: Principles, methods and applications’. *Knowledge Engineering Review* 11, pp. 93–136.

Wada, K., Shibata, T. (2007). ‘Living with seal robots: Its sociopsychological and physiological influences on the elderly at a care house’. *IEEE Transactions on Robotics,* 23 (5), pp. 972–980.