VMPepper: How to Use a Social Humanoid Robot for Interactive Voice Messaging

Paola Barra
Dept. of Computer Science
University of Salerno
Salerno, Italy
email:pbarra@unisa.it

Carmen Bisogni
Dept. of Computer Science
University of Salerno
Salerno, Italy
email:cbisogni@unisa.it

Riccardo Distasi
Dept. of Computer Science
University of Salerno
Salerno, Italy
email:ricdis@unisa.it

Antonio Rapuano
Dept. of Computer Science
University of Salerno
Salerno, Italy
email:arapuano@unisa.it

***Abstract*—VMPepper is an encrypted voice mail system managed by Pepper, the humanoid robot. Pepper ships with face recognition software that has been used in voice mail system design. The robot has a voice recording module as well. A human that has already been recognized can record or listen to messages to or from another user. Sensitive data such the audio message itself, sender and recipient are encrypted during transmission. A set of users has been selected to evaluate the system. During the experiments, Pepper could move freely, initiating interaction on recognizing a user. The interactions completed successfully in 100% of the cases.**

***Keywords*—** *Humanoid Robot; Pepper; Face Recognition; Voice Mail; Cryptography; Cloud.*

I. Introduction

Humanoid robots are evolving and becoming increasingly simple to use and program. At the same time, sensors are getting less and less invasive. This progress makes it possible to design helper applications that interact with humans in a more casual and relaxed way. The particular humanoid robot used in this project is Pepper, designed and developed by SoftBank Robotics Corp. and Aldebaran Robotics SAS [2]–[4]. It is depicted in Figure 1.

Pepper is not the first robot produced by Robotic SoftBank: there is a direct predecessor named NAO, which also runs the NAOqi operating system. The simplest way to develop custom applications is the box programming environment Choreographe, available on all of Aldebaran's products. Snippets of code are pasted into text boxes that get activated under specific circumstances. Choreographe's SDK can interface with several programming languages; the choice for this project was Python 2.7.

Both Pepper and NAO are suitable for interaction with humans [5], [17]. In particular, Pepper presents itself with a childlike appearance due to its height of 120 cm, its large eyes and other soft facial features. This eases humans into a more spontaneous and cooperative interaction. Pepper, in fact, has already been used for experiments and shown to facilitate people's existence, as reported in [1]. Pepper's main specifications are shown in Table I.

The paper is organized as follows. The next section outlines the state of the art regarding Pepper and other humanoid robots used in a variety of applications, as well as non-humanoid robots specifically offering answering machine services. Section III describes the method used for the present proposal, while Section IV illustrates the experiments involving interaction with the robot. Finally, Section V presents our conclusions and possible future developments.

II. Previous Work

Previous research work using Pepper and other humanoid robots has been centered on three main areas: interaction with the environment, medical applications, and social interaction including education. In all three cases, the sensor component is the main driver

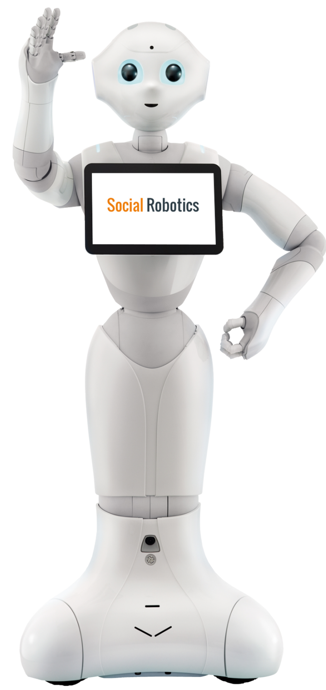

Fig. 1: The humanoid robot Pepper

TABLE I: SPECIFICATIONS

Hardware and Connections	Details
Size (H x D x W)	121 x 425 x 485 [mm]
Weight	28kg
Battery	Li-ion 30.0Ah / 795Wh
	3D sensor × 1, touch sensor × 3
Sensors (trunk)	Gyroscope sensor × 1
Sensors (hand)	Touch sensor × 2
Sensors (leg)	Ultrasonic sensor × 2, laser sensor × 6, bumper sensor × 3, gyroscope sensor × 1
DOF	20
Display	10.1 inch touch screen
OS	NAOqi OS
Network	Wireless and wired interfaces
Speed	Max. 3 km/h

for innovative use.

1) Interaction with the environment: Allowing Pepper to walk inside a closed building is a challenge because GPS sensors cannot be used. The most natural solution is try to make the most of the information provided by other sensors such as cameras and proximity sensors. In [6], a 3D map is created so the robot is able to move independently, avoiding obstacles. In [7], the indoor trajectory that the robot will follow is calculated a priori. These papers show that it is possible to create “intelligent space” with just visual sensors. On the other hand, the content in [8] and [9] covers more technical details, in particular a study of the maximum inclination that the robot motors can withstand. Management of physical contacts with the robot is also investigated through the use of proprioceptive sensors.

2) Medical care: Humanoid robots can be useful in the medical field, too: for example to manage anxious patients as shown in [10]. Another possible application is as a medical assistant: the robot is used for daily patient data collection [11], or to help patients respect their prescription drug schedule [12]. By charging the robot with work previously entrusted to man, human staff gain time while patients manage to deal with robots in the simplest tasks.

3) Social interaction and education: Social interaction with Pepper is widely used to offer services and entertainment. When using Pepper in human interaction, it is necessary to use sensors to collect feedback and human emotions. A robot programmed for social entertainment typically interacts based on its own perception of human emotions, and can respond to such emotional stimuli by showing emotions itself, as shown in [15]. In other experiments, a similar approach was followed to interact with children, and Pepper turned out to be an effective interactive educator [16].

Applications closer to our voice mail system are provided by well known commercial voice assistants such as Alexa, Siri, Cortana, and the like. A detailed description of the services offered by these systems can be found in [18]. The interaction in these cases is exclusively vocal, so a point by point comparison with Pepper is not possible. However, the services offered by a typical voice assistant are quite similar: making phone calls and sending or reading text messages and emails. The hardware in these commercial systems does not include a range of sensors as wide as Pepper’s, so face recognition is out of the question. On the other hand, the variety of motors, motion sensors and cameras offered by Pepper enable free range movement and face recognition to be an integral part of the services offered.

III. METHOD

The proposed system is an interactive service. The general design goal is that the users should be able to interact with the robot as if it were an intelligent answering machine. After recognizing the user, the robot records a message for a specific recipient, who should also be registered with the system; when the robot meets a recipient in its pending message list, it asks them if they want to listen to the message or record a new one. Data privacy is based on facial recognition. In fact, facial recognition is performed twice: when a user approaches the robot for the first time, and when a message is to be recorded.

Voice Mail Services offered by the robot are restricted to authorized users. Therefore, a registration phase is required. The registration process needs help from a human operator. The operator starts recording face features with the “Learn Face” box of Choregraphe. The extracted features are stored in robot memory, and they are recalled during the face recognition step, an example of which is illustrated in Figure 2.

A. Interacting with Pepper

Users that want to interact with the robot must approach it in order to get recognized and therefore authenticated as a registered user. When the robot recognizes a user, it acknowledges them by saying ”Hello *Name*”, as shown in Figure 3.

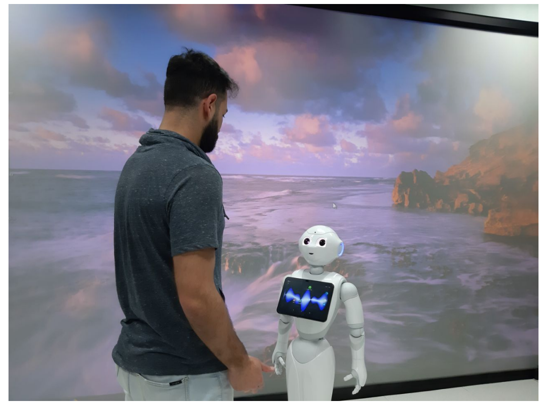

Fig. 2: Pepper learns a face

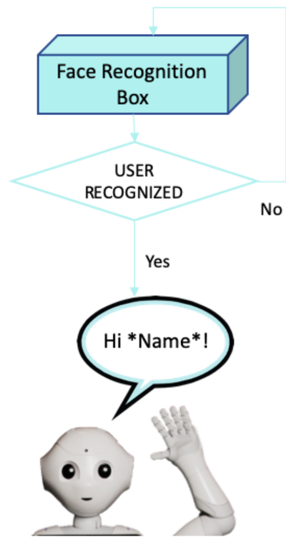

Fig. 3: Pepper recognizes a user

The user can access the voice mail services by uttering phrases with keywords “Leave” or “Listen”.

If a phrase with the keyword “Leave” is pronounced, the interaction proceeds as follows.

1) The robot asks “Who is the message recipient?”.
2) If the name the user says is not stored in memory, the robot will reply “Im sorry, your friend is not registered”.
3) Otherwise, the recipient is well defined, so Pepper notifies the sender that recording is starting.
4) Once recording is done, Pepper tells the sender that the message will be delivered—that is, replayed—to the recipient as soon as the occasion arises—that is, as soon as Pepper meets the recipient and completes facial recognition.
5) The process returns to the face recognition step.

If a phrase with the keyword “Listen” is pronounced, the interaction proceeds as follows.

1) The robot asks the user “Who is the message sender?”.
2) If there is no match between sender and recipient in the stored data, the robot says ”Sorry, there are no messages for you from him/her”.
3) If there is a match, Pepper performs face recognition again. The face recognition at this step is to enforce basic privacy/ security: if the would-be recipient moves away from Pepper,

 ISBN: 978-1-61208-724-5

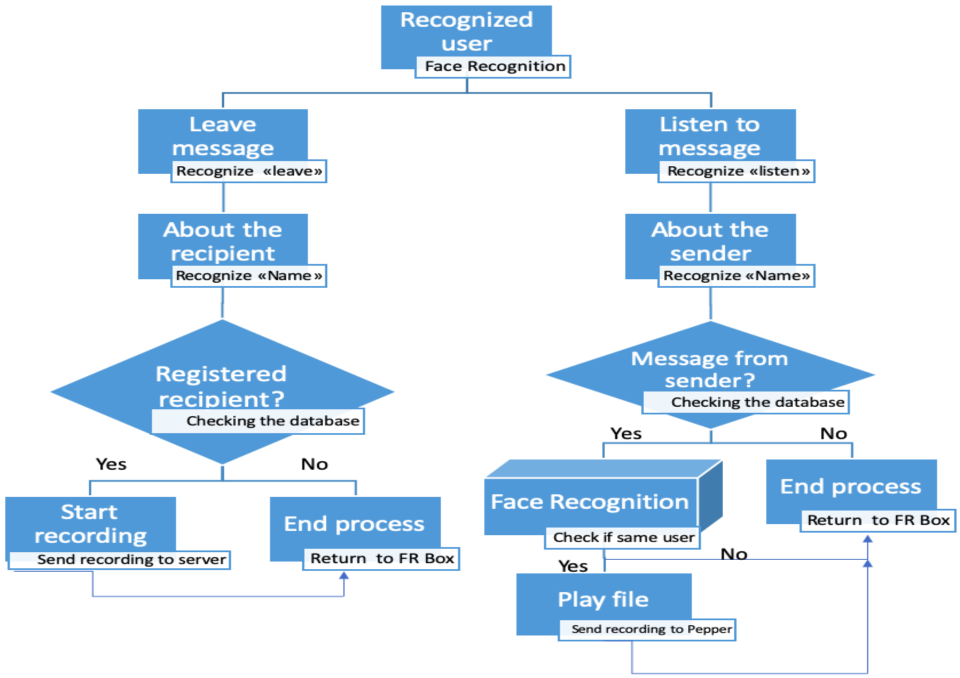

Fig. 4: Proposed workflow

or if some other user tries to take over the interaction, the message will not be played.

4) If face recognition succeeds, the message is played.
5) The process returns to the face recognition step.

Both procedures are visually summarized in Figure 4.

B. Pepper-Server data flow

The voice mail method system is made by two subsystems, both of which rely on NAOqi's library to exploit Pepper's capabilities. A remote server is used to perform part of the tasks.

- The *Record* subsystem, related to the "Leave a message" module, stored and executed on Pepper.
- The *Replay* subsystem, which performs audio file upload and runs the "Listen to a message" module, stored and executed on the server.

The first subsystem runs the face recognition module. After the interacting user has been recognized, Pepper will listen for the keywords "Leave a message" or "Listen to a message".

When the "Leave a message" module is activated, the Record subsystem, running locally on Pepper's operating system, asks for the recipient of the message. If the recipient is recognized as a registered user, a voice message is recorded. After that, the Record subsystem sends an encrypted HTTP request to the Replay subsystem, running remotely. The parameters of this HTTP request are the sender's ID, the recipient's ID, and the audio file.

When the request is processed successfully, the Record system will delete the audio file from Pepper's local memory. The Replay subsystem runs on the server and listens for HTTP requests. As soon as a request arrives, it stores the audio file on a cloud storage service (Google Drive in the first prototype) and adds a new item into an associative array. The item contains the sender's ID, the recipient's ID, and a link to the audio file on cloud storage.

When Pepper recognizes the "Listen" keyword, it sends a message to the Replay subsystem, which in turn checks for the listener's and the sender's registration. If both are in the list, Pepper performs face recognition on the listener. Finally, Pepper's "open WebView" module is run with the cloud storage audio link as a parameter, so the audio is actually played out. Processes are shown in Figure 5.

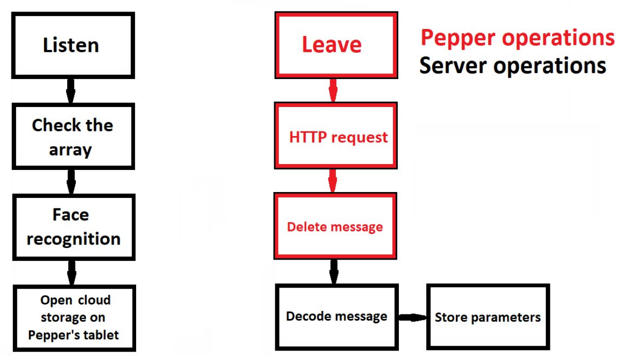

Fig. 5: Pepper-Server local-remote data flow

IV. EXPERIMENTS

In human-robot interaction, the qualitative indexes try to assess the user's feelings during the interaction, as well as the robot's ability to conclude the interaction according to user expectations.

Pepper has an operating parameter "autonomous life" that can be switched on or off. The parameter dictates if the robot can react to external stimuli besides those managed by its current custom programmed behavior. Therefore, when autonomous life is off the robot cannot be distracted by sounds in the background—or even by its own limb movements. When autonomous life is on, the robot reacts to the external stimuli according to its factory settings. Therefore, with autonomous life on, the robot is able to reply to user questions, turn its head to follow a locked-in user face, or turn its body to point the camera in a suitable direction: as an example, if a sound is heard, Pepper will try to aim the camera to the apparent source of the sound.

Leaving autonomous life set to on means that Pepper's behavior is more natural. The autonomous movements make for better and more fluid interactivity. However, there is the tradeoff of external interference that could make interactions harder to program and to carry out in practice. It is not even necessary to have a malicious agent trying to disrupt operation: random external events might be quite enough.

For this reason, two sets of experiments were performed, with autonomous life respectively off and on.

A. Autonomous life off

The user pool consisted of five people of age between 23 and 30. Their faces were recognized as described at the beginning of Section III, and they were added to the registered user set. They were provided a brief verbal explanation about Pepper's services and voice commands.

Each user in this block of experiment interacted with Pepper 10 times, for a total of fifty social interactions. We set a maximum message duration of 10 seconds, so that they could leave meaningful messages without having to wait for longer timeouts and without the need to program an "end message" aural signal.

The results are quite encouraging. Pepper carried out face recognition and correctly recorded and forwarded the message in 100% of cases. Furthermore, the users found the workflow of method very natural and friendly.

B. Autonomous life on

With autonomous life turned on, the experiments followed a different protocol.

- We set the robot free to move in a room in any direction, performing all possible translations and rotations.
- When the robot sees a user that it can recognize, it stops.
- It says hello to the user and starts the flow as described above.

In this operating mode, we found that it is difficult for Pepper to recognize users if they are moving. This was expected. However, if the users exploit Pepper's autonomous life to divert its attention to themselves by making suitable sounds, the robot turns its head and recognition happens easily. The flow then proceeds in the same way as in the "autonomous life off" mode.

V. Conclusions and Future Work

We observed that the use of a humanoid social robot as a voice messenger is quite natural and enjoyable for the user. The ability of a humanoid robot to move autonomously in large spaces makes this application useful for a large population of users. Interaction is efficient and without significant glitches, particularly with the face recognition box and message recording. This may enable extended applications, particularly in combination with other services. As an example, if a mapping of the entire building or complex is available, the robot may be able to bring the message to the recipients directly rather than waiting to meet them.

References

[1] Amit Kumar Pandey and Rodolphe Gelin, Pepper: "A Mass-Produced Sociable Humanoid Robot: Pepper: The First Machine of Its Kind", IEEE Robotics & Automation Magazine, Volume: 25 , Issue: 3 , Sept. 2018, pp. 40 - 48.
DOI: 10.1109/MRA.2018.2833157

[2] CNN. Meet Pepper, the emotional robot. Retrieved February 24, 2015, from https://edition.cnn.com/2014/06/06/tech/innovation/pepper-robot-emotions/index.html

[3] TIME. Meet Pepper, the Robot Who Can Read Your Emotions. Retrieved February 24, 2015, from https://time.com/2845040/robotemotions-pepper-softbank/

[4] IEEE SPECTRUM. How Aldebaran Robotics Built Its Friendly Humanoid Robot, Pepper. Retrieved February 24, 2015, from https://spectrum.ieee.org/robotics/home-robots/how-aldebaran-robotics-built-its-friendly-humanoid-robot-pepper

[5] Arkadiusz Gardecki and Michal Podpora, "Experience from the operation of the Pepper humanoid robots", 2017 Progress in Applied Electrical Engineering (PAEE), 15 August 2017.
DOI: 10.1109/PAEE.2017.8008994

[6] Eiji Kaneko and Nobuyuki Umezu, "Rapid Construction of Coarse Indoor Map for Mobile Robots", 2017 IEEE 6th Global Conference on Consumer Electronics (GCCE 2017), 21 December 2017.
DOI: 10.1109/GCCE.2017.8229367

[7] Dorota Belanová, Marián Mach, Peter Sinčák and Kaori Yoshida, "Path Planning on Robot Based on D* Lite Algorithm", 2018 World Symposium on Digital Intelligence for Systems and Machines (DISA), August 2018.
DOI: 10.1109/DISA.2018.8490605

[8] Jory Lafaye, Cyrille Collette and Pierre-Brice Wieber, "Model predictive control for tilt recovery of an omnidirectional wheeled humanoid robot", 2015 IEEE International Conference on Robotics and Automation (ICRA), 02 July 2015.
DOI: 10.1109/ICRA.2015.7139914

[9] Anastasia Bolotnikova, Sbastien Courtois and Abderrahmane Kheddar, "Contact Observer for Humanoid Robot Pepper based on Tracking Joint Position Discrepancies", 2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), 08 November 2018.
DOI: 10.1109/ROMAN.2018.8525774

[10] Sachie Yamada, Tatsuya Nomura and Takayuki Kanda, "Healthcare Support by a Humanoid Robot", 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), 25 March 2019.
DOI: 10.1109/HRI.2019.8673072

[11] Daisy van der Putte, Roel Boumans, Mark Neerincx, Marcel Olde Rikkert and Marleen de Mul, "A Social Robot for Autonomous Health Data Acquisition Among Hospitalized Patients: An Exploratory Field Study", 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), 25 March 2019.
DOI: 10.1109/HRI.2019.8673280

[12] Keitaro Ishiguro, Saki Minamino, Jun Kawahara and Yukie Majima, "Development of a Robot Intervention Program in Medication Instruction at a Pharmacy", 2018 7th International Congress on Advanced Applied Informatics (IIAI-AAI), 18 April 2019.
DOI: 10.1109/IIAI-AAI.2018.00198

[13] Chiao-Yu Yang, Ming-Jen Lu, Shih-Huan Tseng and Li-Chen Fu, "A companion robot for daily care of elders based on homeostasis", 2017 56th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE), 13 November 2017.
DOI: 10.23919/SICE.2017.8105748

[14] Thi Le Quyen Dang, Nguyen Tan Viet Tuyen, Sungmoon Jeong and Nak Young Chong, "Encoding cultures in robot emotion representation", 2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), 14 December 2017.
DOI: 10.1109/ROMAN.2017.8172356

[15] Wen-Feng Shih, Keitaro Naruse and Shih-Hung Wu, Implement human-robot interaction via robot's emotion model, 2017 IEEE 8th International Conference on Awareness Science and Technology (iCAST), 15 January 2018.
DOI: 10.1109/ICAwST.2017.8256522

[16] Fumihide Tanaka, Kyosuke Isshiki, Fumiki Takahashi, Manabu Uekusa, Rumiko Sei and Kaname Hayashi, "Pepper learns together with children: Development of an educational application", 2015 IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids), 28 December 2015.
DOI: 10.1109/HUMANOIDS.2015.7363546

[17] Softbank Robotics Documentation Retrieved 26 October 2018 from http://doc.aldebaran.com/2-5/index.html

[18] Matthew B. Hoy, "Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants", Medical Reference Services Quarterly, Volume 37, 2018 - Issue 1, pp 81-88.
DOI:10.1080/02763869.2018.1404391