Pi²: PRESENTATION & SUBJECT ANALYSIS

Identification :

Project Title	InMoov : Robot Intéractif
Project Number	66
Team Referent	Corentin LEMAITRE
Other members in team	Rabah HOUAOUI ; Kristof SZUCS ; Rayan GALETTA
Team Mentor	
Project Partner	DaVinciBot

Context

Nowadays robotics is a science that is beginning to take an important place in our daily lives. Robotics is the science interested in robots, it's an interdisciplinary branch of engineering and science which includes mechanical engineering, electrical engineering, computer science and more particularly Artificial Intelligence. There are different branches of Robotics, as part of our Pi² project we're going to focus on humanoids robots. A humanoid robot is a robot whose general appearance recalls a human being. Generally, humanoids robots have a torso with a head, two arms and two legs. Some humanoids robots may have a face with eyes and a mouth. InMoov fits into this branch of robotics.

In 2012, Gaël Langevin a French sculptor and designer gave birth to the first open source humanoid robot. This human sized robot is fully printed in 3D. It has been designed so that it can be printed on any printer with a print area larger than 120*120*120 mm.

The fact that InMoov is open source implies that the robot is continuously evolving thanks to the contribution of its very active community. The project started from the will of Gaël Langevin to create a prosthetic arm. He then started uploading the files on the internet for the community. Following the positive reaction of this initiative, he decided to embark on a complete robot.

We have chosen this robot among others because it contains a very strong creative potential. InMoov is a modular robot with many characteristics, we can cite as examples:

- Its structure is entirely printed in 3D. It's very simple to replace, customize or even improve certain parts.
- Its large number of sensors: The possibility of adding new sensors to those already existing, opens the way to innovations on the robot electronics and on interaction with his own environment.
- The open source aspect is a high added value to the project because we would like to fully contribute to the evolution of InMoov and make the whole community benefit.

First State of the Art

InMoov's State Of Art

There is a few software available for the control of the robot but two of them show up more than others:

- **MyRobotLab (MRL):** It's an open source service based robotics framework written in Java which has bindings for Python. It also has a Web UI written in AngularJS which allows remote control. It's adopted by the largest part of the InMoov community since it supports InMoov directly. If we you use this software you must use NervoBoard and Arduino microcontrollers. MRL also allows us to use virtual InMoov which can be helpful to develop or test without the physical robot. The only issue is that it's hard to find good information on the software and MRL is hard to start with.
- **EZ-Robot:** Uses other microcontrollers that can run any hardware (camera, servos, sensors). It's more user-friendly but doesn't use the same microcontroller as the creator of InMoov. It's not meant to be for developers, more for people that don't want to program.

For the interaction, Gaël's InMoov uses:

- For Text To Speech engine, MarySpeech, Polly, NaturalReaderSpeech or VoiceRss. AcapelaSpeech & GoogleSpeech are also used sometimes but they need internet connection. He can speak 6 languages: English, French, Spanish, Russian, German and Dutch.
- For a chatbot: Program AB with AIML files.
- For mouth synchronisation: Mouthcontrol activates a software jaw movement while the robot speaks or Audio signal processing which makes real-time jaw movement based on audio signal.

Other interactive robots

Pepper (SoftBank Robotics)

Pepper is a human-shaped companion robot able to recognize facial expressions and human emotions. We can talk with Pepper in more than 20 languages fluidly and intuitively. Depending on the user's mood and expressions Pepper can adopt its behaviour and tone. The user can personalize the robot and the robot itself can automatically learn and adapt to its user's personality. Pepper's emotions are expressed through a tablet attached to its chest and not on its face. Pepper can blink by turning off the light above its eyes which makes the robot more realistic. This is one of the few robots that is able to dance on the rhythm of the music it plays. Pepper is already in service in more than 140 mobile stores in Japan and trialled at Belgian hospitals. The robot's main task is to welcome, inform and amuse the interlocutor.

Nao (SoftBank Robotics)

Nao is a 58cm tall humanoid customizable companion robot. Nao can provide reception and concierge services. He welcomes the guests and provides them with information in many languages. Nao is also used to help and interact with autistic people. The user is able to learn and have trainings related to its needs and skills. Nao is also able to learn from its user but we still can't talk about real artificial intelligence. Nao show a significant capability of precise motion control, however walking speed is not as fast as we would expect. The user can program unique behaviours for Nao using Python, C++, Java, JavaScript.

Buddy (Blue Frog Robotics)

Buddy is a half meter tall companion robot. Buddy's head includes a tablet that works with several sensors built in its head and body. The robot can easily move around in a household on its 3 wheels without any object collision and help the users in many different activities. The robot is able to communicate in plentiful languages, work as personal assistant, receive messages, phone or video calls, play music from our favourite playlist or inform us about weather or traffic jams. Buddy can recognize and follow humans face or voice and also objects. Buddy can be remote controlled and patrol our home with its camera detecting any

unusual activities. This little robot can also interact with most of the Wi-Fi connected objects in the house. Buddy can help kids learning or elderly remember taking medication. Buddy is based on an open-source platform using popular development tools like Arduino, OpenCV and Unity3D.

Jibo

Jibo is a 30cm companion robot with a simplistic design. The robot is stationary but able to turn around in 360°. Jibo is equipped with two cameras, microphone, stereo and a touch screen. Jibo can recognize and follow people's face and voice and communicate with them. Wi-Fi connected objects can be connected to Jibo. This home robot is able to learn from conversations and its experiences. Jibo can work as a personal assistant handling mails, sending reminders, taking photos or making video calls. We can name three key principles concerning Jibo's behavior: starting conversation when a person is recognized, individualizing the conversations and interacting with people in a human-manner.

Promobot:

Promobot is a Russian humanoid robot on wheels, serving promotional and advertising functions. The head is equipped with a display that can show the robots' emotions or facial reactions. On the chest of the robot, we find a touch screen showing related information about the customers' needs. The robot was made to work in areas with a high quantity of people. Promobot is able to communicate with its users and navigate them precisely to any destinations. The robot can automate the registration process of visitors and share information with them about sales and good plans. Promobot remembers all the people who talked with it and (in theory), increases consumers purchase intentions and loyalty. Promobot collects customers' feedback as a data that can be later analysed by the owner.

Different forms of interactivity already existing:

We have seen previously that there are many humanoid robot, each with its own characteristics but also with a different interactivity.

First of all, we can think of the most spontaneous interactivity, the Man to Machine talk. The aim of this interaction is to have the most natural dialogue with the robot, for example when we question him on his state of mind or request him an information.

The second form of interaction which is not used by a lot humanoid robot is the sign language. In a general way, sign language requires the use of one's arms, hands, fingers and facial expressions. For a humanoid robot to master this form of interactivity implies at first a big coordination between his arms, his hands, his fingers but also, his face. Moreover, since the sign languages can be different between each country and area, it's very difficult to learn a robot to recognise it. The number of humanoid robots able to master this form of interaction is very small. The most recognized robot in this field is Nino, a humanoid robot designed by Taiwan researchers and students.
In recent years, some humanoid robots have been redesigned with a tactile tablet on the chest. The purpose of the touch pad is to significantly increase the functionality of the robot. With this tool, we can ask the robot to do a research (get a restaurant address, send an e-mail, know the weather of the day, ...) with a simple interaction and the robot will provide vocally the information collected. Pepper robot uses this technology, but the newest robot that uses this feature is Reem a humanoid robot policeman. This robot can collect reports of crimes or misdemeanours with his touchpad.
Moreover, a chatbot is a form of interaction that is beginning to spread everywhere, it is a program designed to simulate a conversation in natural language using auditory or textual methods. We can find this form of interaction more and more on websites that appears in the form of an animated character.

We've just discussed forms of interaction that are strongly related to language, but there are also interactions that are not based on language but on gestures and this is what we are going to deal with in this section.

Some robots are able to grab objects and bring them to a specific place. For example, you can ask the robot to get the coffee. However very few robots are able to perform this kind of action, the most complete humanoid robot in this field remains ASIMO a humanoid robot designed by the company HONDA.
Through this state of the art we have seen that there are different forms of interactivity, now we have to find one or reuse one in order to develop our own.

Objectives and sub-objectives definitions

As speech is the most effective and natural way to interact between humans, InMoov wants to perform in this discipline by being able to communicate without any tools beside himself. Conversing with machines is a long-standing dream of human-computer interaction, but in the last few years their ability to understand natural speech has been revolutionized by the application of deep neural networks.

1. A robot capable of listening. Speech Recognition

The three steps towards a discussion with the robot, is to able him to recognizes our sentences.

How to hear?

The first step is to define the nature of the context that will lead the user to use InMoov. In our case, InMoov has to be able to communicate with anybody, and on context blurrier than in a quiet room. This is why, we must avoid software like Dragon that rely on machine learning of a single user, but instead turn to multi-locution software. As we still need to keep a decently low ratio of mistakes, using Google Speech Recognition to transcribe the user's speech to text seems like a good idea, as it relies on an enormous cloud library grammar.

How to analyses those sentences?

Once the words are captured by the microphone and the previous software, the transcripted text needs to be interpreted to create a meaningful sentence. Meaning that InMoov should be able to differentiate spoken words from noisy events, and put them together to form the closest sentences. This is where tools like jasper, Webkitspeachreconition, or Microsoft.speech appears on screen.
Again, as different techniques exist the objective must be redefined. Some software relies on specific keywords to answer, (e.g. User: Hello, could you give me the weather of Paris Software: *weather Paris*). Such a technique would improve the efficiency to do actions but drastically cut his ability to answer with so little left of the sentence. This is why, in order to keep enough information to answer back, we need to use a software powerful enough to fill in the words not understood with the meaning of the whole sentence!
The software that we will most likely have to use is the Google Speech API and here is why. We must work knowing that every software and hardware that we choose has to be compatible with the robot structure and useful to the other team! So, a lot of component such as Arduino card with specific software goes way off our goal. The handy thing from this API is that it processes the data recorded in the cloud with dozens of functions, and return the text that we will need but also tons of information related to the emotion that would have been perceived by a human. These last criteria are essential as the "Robot Emotif" team also need those information from the vocal communication.

2. How to make InMoov answer back Speech

Once that the conversion Speech to text has been made, two extra steps have to take place before hearing InMoov talk.

Building sentences

Another ground-breaking technology just reached levels where it can be successfully used by any company (and us) to precisely imitate human's behaviour on various scenario, Chatbots. Our goal is to use a chatbot powerful enough to process natural language yet functional in order to program certain task. As a Chatbot intelligence is measured by its capacity to handle any scenario and to be aware of the user need, we

will most likely be using a cross between two. Cross to take out the best from bots that specialize in talking, and another Bot acting like a personal assistant. Mitsuki appear to be the most human-like chat bot available but dissociate itself from the regular personal assistant that every smartphone seems to have. On the other hand, open source chatbot within MyRobotLab tend to have a good balance between communication and execution of precise task.

Speaking

We want the users to be able to maintain a casual conversation with InMoov, however a robotic, sharp and stoic voice would simply drive away the attention from the content to the mispronounced sentence. In order to voice those text, we are going to use a TTS (text to speech). Nowadays, 2 different ways have reached the global community. The concatenative TTS where a very large database of short speech fragments is recorded from a single speaker (SIRI) and then recombined to form complete utterances. Or the parametric TTS, where all the information required to generate the data is stored in the parameters of the model, and the contents and characteristics of the speech can be controlled via the inputs to the model.
The first one is the most used, as it is more stable and generally simpler, but a new way is emerging and burst every previous attempt, WaveNet. By directly modelling the raw waveform of the audio signal, one sample at a time, it renders a more natural-sounding.
So, after comparing this technology with the material available for our InMoov we will have a closer look on which approach will please our ears the most!

3. An intelligent robot

The main purpose of having a great human-robot interaction is to help humans. For this sake, the robot has to bring around his knowledge, in the most logical and consistent way, and to be useful, this knowledge has to be improved.

Coherence of his conversations

Our humanoid must be able to interact we users with coherence.
First of all, he should be able to answer a question by staying in the context of the whole conversation, he should reply in harmony of the last exchanges.
Secondly, if the robot is having a conversation with multiple users, he should be able to notice who is speaking. The aim of this objective is to let him access the global context of the conversation.
However, if the robot is not sure of his answer, he mustn't share it. Instead of giving possible wrong answers, he should just tell the user he doesn't know the solution to the question, or doesn't know what do. Otherwise, if you make him react, even if he's not sure that's the right move or reply, the user will often receive wrong response.

Improved knowledge

For a simple question our InMoov should be able to reply with his own database.
In order to be clever enough to answer difficult question, the robot must be able to access data. We'd like to link his AI to different data sources that could increase his knowledge on which he can rely on in case he's having trouble satisfying the user. For example, we can use other databases, APIs, or screen scrapping.
To reply the quickest way possible, these data sources must be efficient, reliable and easy to access to. The aim is to answer the best and fastest way possible.

Growth of his knowledge and Machine learning

In the interest of having a robot that can learn fast and easily we would like to put together his previous knowledge, and experience, with the new things that he can find out. He should, with a particular instruction, be able to get in "Learning Mode". This option would let us teach him, give him some new information or update obsolete information.
He must have the means to make faster response to the ones already asked by a user. He should be able to use previous request data stored in a kind of cache so that future requests can be served faster. It would also be great if he could learn by himself, making updates on daily data that are often requested by the users.

Valuing

What is valuing your project?

Artificial Intelligence is a vast subject that has always made people fantasize. Today, a new industrial era is opening before us where human cognitive functions are in competition with the machine. If the stakes of industries are colossal, it seems essential that the training of the engineers of today is in adequacy with the needs and the questions of tomorrow. InMoov, is the perfect combination of what the world will most likely need in a decade. Especially for us, the interaction teams, where developing recognition program and chatbot AI is relevant as most of the leading tech company (cf Microsoft experience 2017) are starting to promote such tools to revolutionize every user experience in app and websites. With our project, we have the opportunity to, not only better our competences, but also, raise new problematics and solutions for the whole InMoov community.

Nowadays, over 3000 members are contributing to this community which is getting bigger every day. Each member is free to publish his new improvement on their robot by sharing it with the open-source community. We would like to share our own upturn related to the InMoov/human interaction amelioration that we can add to the basic InMoov.
For example, we could use this interaction for school demonstration, (e.g. Open doors days, Associative presentation...). Also, all this innovation must be usable by the next generation that would like to keep going deeper into our work and push its limits.

How can the valuing be complete?

To fulfil the project, this robot should serve a clear purpose. As future engineers, it is important to keep a fair balance between feasible and needed. For InMoov, the balance should be to make a top tech robot yet user friendly.

We would like to use different APIs to achieve a great human/robot interaction. First of all, to be able understand the speech of the user we need to implement a speech recognition API that will translate the sound received by our InMoov into a text. This text will be analysed to find out the right way to reply to the user. In order to retrieve the correct information needed to answer, we'll use different online APIs and our own database depending on the nature of the question.

To maintain a natural conversation with the robot, it has to be the most human like. Which is why InMoov needs to give the most proportionated answer in terms of information, volume and tone. To be able to give a quick respond, whether or not the answer is known, the robot must be powerful enough to process it all and create a meaningful sentence while being aware of its own capacity.

One of the innovations that makes our robot different from the already existing robot is the knowledge and database about the Pôle Leonard de Vinci that it could use. As already mentioned, we aim to use our InMoov during Leonard de Vinci Open Days to promote the schools. The robot will attract students who are interested by new technologies and robotics. Their interest toward our project would give us the opportunity to present the schools and make an impression on the students. They will be able to ask precise questions about lectures, associations, future job options and have an adequate and direct answer from the robot.

Eventually the robot could be used to register those students' name, e-mail and other information by simple conversations. The registration process could be speed up by InMoov or if not, at least it could be more interesting for the students. The data could be added to the database and treated after like any regular student registration.