Implementation of a Virtual Coach for Weightlifting Athletes with MARC Virtual Agent Platform

Angelica Acosta Arteta angelica.acosta@telecom-sudparis.eu

Daniel Rodríguez daniel.rodriguez@telecom-sudparis.eu

Evelyn Bankell evelyn.bankell@telecom-sudparis.eu

ABSTRACT

The development of technology is progressing and we humans are searching for new technical tools to interact with to make everyday life easier. An interactive virtual character has been demonstrated to be able to guide and help people likewise as a human being. An important role for how humans perceive information is through emotions and expressions, especially facial expressions. This paper presents a system for an interactive virtual character that coaches an user through different workout sessions using a real-time multimodal communication. The virtual coach is using facial expressions such as happiness for indicating a good executed job and to increase the engagement of the user. As well as a displeased facial expression for indicating a job not executed as expected and to give the user a feeling of disappointment. This paper presents a software architecture using the MARC virtual agent platform and how a state machine is implemented followed with modules and design considerations.

Author Keywords

Emotions; sports; coach; virtual human;

INTRODUCTION

In the present day virtual coaches are becoming more popular providing support for different activities that require human interaction, for example, therapies, rehabilitations and behavior change [6]. All of these activities involve emotional responses because of the human factor presence in the interaction. Humans are capable of encoding and decoding emotions and provide a proper feedback, according to the emotion inferred [2]. Therefore, the recreation of emotions by a virtual coach is the core of the development of these types of tools. The virtual coach must be capable of identifying an emotion and provide a proper feedback because the principal objective is to replace the human factor and virtually recreate a human-human interaction.

In the first section, we present the ideas that serve as a ground work of our investigation; commun emotions in sports, emotions representation in computer systems and how they are inferred. In the following section, we provide an illustrative scenario to give an example of a situation where the virtual coach can be used. Afterward, the architecture of the system is described, including the MARC modules functions in this architecture. In the fourth section, a set of mock ups are provided to illustrate the graphical user interface of the system. The fifth section explains the implementation of the virtual coach with a MARC state machine, including explanations of different situations that can be handled by the virtual coach. The sixth section presents the results obtained by testing this new system. Finally, we provide some insights comparing the methods used to implement the virtual coach presented here and other methods.

RELATED WORK

Physical Activity Recommendations

Physical activity is a key factor in achieving a good and healthy lifestyle. It has been shown that following appropriate levels of physical activity improves muscular and cardiorespiratory fitness; reduces the risk of hypertension, stroke, diabetes, various types of cancer, and depression; as well as being a way of energy consumption which helps with energy balance and weight control [8].

The World Health Organization recommends that children from 5-17 years old participate in at least 60 minutes of moderate-to-vigorous levels of physical activity daily; adults aged 18 years and above participate at least in 150 minutes of moderate-intensity aerobic physical activity per week; with special recommendations for the ones aged 65 years and above, as this group tends to be the least physically active. These recommendations, however, must be adapted to the current activity level of the individuals [8].

Encourage Physical Activity Through a System

Motivating individuals to accept interventions encouraging physical activity behaviors through persuasive strategies is a possible option that has received a lot of interest in the research community. Petsani, Kostantinidis, and Bamidis [3], propose the following Behavior Change Technique for an e-coaching system that consists of 3 phases: the starting, the design and adaptation and the intervention phase.

In the starting phase the main focus is to find the participant’s motivation to start with the activity in order to assess and enhance it, by explaining the benefits and proposing evidence that the participant could reach its goal.

In the design phase occurs the personalization of the intervention. Here the system will propose some activities based on the participant’s abilities, to make sure that they are realistic and not overestimating. It is really important that the participant helps choose the tasks, as this will help to engage him in the process.

In the intervention phase takes place the tasks that were accepted in the previous phase, and the aim is to keep the interest by rewarding the participant when achieving goals and creating expectation for the future rewards.

The system must learn from the user in order to not lose its trust. If a suggestion is rejected by the participant, the system must ask questions to identify the reasons for its rejection. This way the system can save this information to not pressure the participant into suggesting the same type of activity in future occasions. At the same time, if an option is accepted and successfully completed, the system encourages and praises the participant [4].

The importance of feedback when engaging in a physical activity or sport was studied by Kettunen, Critchley and Kari [1]. They explore how a digital coach can offer guidance and instructions on how to improve training among physically active people.

The study was made on 25 professional cross-country skiers, where the goal was to help the skiers prepare themselves for a ski marathon race. The digital coach in this case includes a smartphone application and an ECG heart rate monitor as a chest belt. The belt also had an accelerometer pod, and data was transferred via Bluetooth. Real-time data was sent to the application and the user received feedback and advice through their connected headphones. Each participant spent eight weeks with the digital coach, four weeks before the race and four after, where the users got a personalized schedule on how to get prepared for the race.

Many of the participants got excited to follow a specific plan, and this feature seemed to be appreciated. Some of the participants were critical of the mobile device and highlighted that it takes time to adjust to a new coach for guidance in the training. However, the real-time feedback was highly appreciated among the users and feedback gave them a different perspective from what they were used to. Another feature that was appreciated was the diversity along the training and different kinds of exercises was suggested than the ones who the participants were used to, this increased the motivation among the users [1].

One of the users got lower motivation because the feedback was too repetitive and wanted to receive more personal and diverse feedback. The most important factors for increasing motivation according to the users was the ability to receive feedback, see development over time and a more variety in the training. However, the feeling that someone always was watching and recording the training, caused stress among some users.

During the race, many participants felt a confidence boost that came from technical improvement and also the positive verbal feedback from the virtual coach [1].

Modeling Emotions

In the sports context the most common emotions are happiness and anger, with evidence of a relationship between coaches' emotional expressions and the performance of their players and the functioning of their teams [7]. In particular there is evidence that coaches' expressions of happiness projected such emotion on the players as well as inference of good performance. This suggests that expressions of happiness and enthusiasm by coaches can promote good sports performance [7].

Since the objective is to positively influence the participants into engaging with the activities, the virtual coach's capability of regulating his emotions is crucial. It is necessary as well to recognize the emotions displayed by the user in order to make the coach react to them adequately. Same principle can be projected for individual coaching as for team coaching, since the coach is influencing independently of the number of players.

Inferring Human Emotions

Tironi et al use the core affect model of Russell to read emotions through their system [4, 6]. The core effect classified the emotions through two values: valence and arousal. Valence is the grade of pleasure or desirability to a given situation and arousal is the grade of excitement of the individual [6]. With these two values they classified the result inside four basic emotions: happiness, sadness, surprise and anger. These emotions represent big categories that can be specialized like Bas R. Steunebrink et al [5] described in the OCC Model.

The architecture proposed by Tironi et al [6] is composed of three parts: the central part that is the Module Manager and two subparts, the UI Manager and the Empathic Module. The UI manager is the representation of a virtual coach to the user, for example, the avatar, the virtual environments, etc. Additionally, we have the Emphatic Module that processes all the input that the user gives through facial expressions and natural language to infer the emotion that the person has. After that the same module can prepare a response accordingly with the inferred emotions and send it to the UI Module to transmit it to the user.

An interesting concept present in the paper of Tironi et al [6] is, the emotions can be inferred also through other ways, apart facial expressions and natural language. They say that the coach can identify emotions via the analysis of the performance of the user in a given physical activity. For example, if a user did well in some activity the emotion that is going to be shown is happiness or if the activity went wrong the user can be sad. The system can use the information of the results and add it to the facial expressions and natural language analysis results to obtain a more accurate result.

In the architecture of the EVC (Emphatic Virtual Coach) Tironi et al [6] define modes of use for the virtual coach. Particularly the following two: one where the coach acts only as an observer, processing all the results that the user obtained in its activities; and another where the direct talk with the coach allows the system to take input through the facial expressions and language used by the user. This is an interesting approach because the virtual coach can have an emotional response even before the user starts talking with the avatar directly, as the system has an emotional response prepared with the statistics of the user in the activities made by him.

Designing the system

In this article, we explain how we designed an e-coach for weightlifting trainees, using a wall display screen. The virtual coach is represented on the screen alongside the image of the user doing the activity in real time (mirror-like). We use the architecture proposed by Tironi et al [6] to identify the emotions of the user while he is doing the training. The coach analyses the results of a particular session to prepare a response to motivate the user to continue the training. The three phases idea proposed by Petsani et al [3] can be applied here. First the coach tries to identify the participant's current motivation levels, then design emotional feedback according to the situation and at last the system makes an intervention talking with the user, suggesting activities and giving positive feedback to keep the participant's motivation. This approach can lead to a more personalized experience that leads to a better reception from the user, as shown in the results of the experiments made by Kettunen E [1].

Limitations

The main problem to consider when designing a system for sport players is that it must be easy to use while doing other tasks, meaning that the communication should be speech-based so the trainee is able to interact without stopping his training session. Another problem to consider is the display of the system, as sports players can not look at a small phone screen while moving and doing other activities. This is why we propose a bigger display (wall-sized).

ILLUSTRATIVE SCENARIO

Stefan walks into his private gym at 2pm and the system turns itself on.
- Sam (Virtual Coach): Hello Stefan, ready to start today's training?
- Stefan (trainee): -nods with his head-.
- Sam: let's start stretching -neutral emotion-.
- Stefan: no, I already did that in the morning.
- Sam: It's really important to stretch right before starting to train. We do not want you to have another injury. Remember last year? -displeased emotion-.
- Stefan: It wasn't a big deal. Let's just start.
- Sam: I'm sorry, but I really don't feel comfortable starting without stretching. Last year you couldn't train for two months after the accident -displeased emotion-.
- Stefan: fine, let's stretch.
- Sam: -happiness emotion-.

The system uses its knowledge on the participant to persuade him into accepting the suggestion. During the session, the system recognizes Stefan's position and makes small corrections about posture during the practice.
- Sam: remember to lean back a little more to distribute the weight -displeased emotion-.

After Stefan corrects his posture, the system rewards him.
- Sam: Yes! Good job. If you keep the good work for next I'll have to text Adrien (Stefan's real life coach) to let him know how much you have progressed -happiness emotion-.

After the session is over and Stefan leaves the room the system says "see you later!" and turns itself off.

SOFTWARE ARCHITECTURE

Base schema

For the design of our virtual coach we used, as a ground work, the paper of Courgeon et al [2] because of its relationship with emotions and how their expressions are displayed using the MARC virtual agent platform. Our design is composed of different modules that make specific jobs.

We also represent the types of input that the user can enter in the system, *See figure 1*. These can be the followings:

- **Voice of the user:** the user can speak to the system to answer the virtual coach's questions. This input is processed by the OCC Module. This module is going to classify the words said by the user, to identify if he is happy, sad, tired or angry. Consequently an emotion is selected accordingly and is provided to the response module.
- **Exercise recognition (Movement sensors):** when the user initiates an exercise, the camera is going to record all the movements to send them to the module of exercise recognition. This module will define if a given movement is correct or not and the response will be sent to the response module.

Each of the inputs are processed in the recognition module. This is one of the core modules of our system. The recognition module is composed of two submodules. The first one is the OCC Module that will receive the voice input of the user. Then this module will process each word to identify keywords that can demonstrate emotions. Hence, this module can infer an emotion and send it to the response unit (module). The second submodule is the exercise recognition module. This module will receive the movements performed by the user and process them. The module will send a binary answer, to specify if the user did an appropriate movement or not. After identifying the correctness it will send an answer to the response module. This answer will be translated into one of two emotions, joy for a good job or displeased for a bad job.

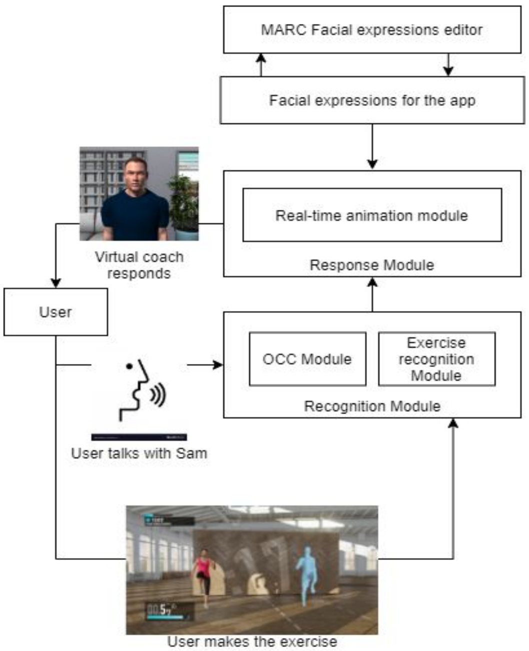

Figure 1. Schema for our Virtual Coach app.

After the recognition module provides an emotion to the response unit, the second module will build an answer to the user. This module applies the tool MARC to build the response, the real-time animation module. With the emotion obtained from the recognition module, the facial expression correlated to the emotion will be retrieved from the database. Afterward, a vocal response is associated with the facial expression and is sent back to the user.

Besides these two big modules is essential to mention the Facial expressions for the app and MARC Facial expressions editor. The first one will contain all the facial expressions necessary for the system to work. It resembles a small database that has responses for the user. Additionally, the MARC Facial expressions editor is used to build the expressions for the system.

MARC modules

The MARC modules displayed in *Figure 2* are a representation of the user input and the output that will be obtained from the coach in different situations. First, the system scans the body and thereby the coach can collect information about the user's posture and position. Otherwise, the system is speech-based and is, therefore, obtaining user input when the user is talking. Thereafter, an automatic lookup via an appraisal database is made and the workout session is starting. As presented in *a*, when the user is following the instructions given by the coach and gets a good result, the coach demonstrates a happy facial expression and delivers positive feedback to the user. However, if the user is not performing correctly, as illustrated in *b*, the coach produces a displeased facial expression. Thus, the coach is not frustrated or angry, instead he presents advice on how to improve. Lastly, in *c*, the module displays an extension from *a*, the difference in feedback occurs when the user follows correctly the instructions in a row. In this context, the coach displays a happy facial expression and gives voice feedback that is more energetic than before.

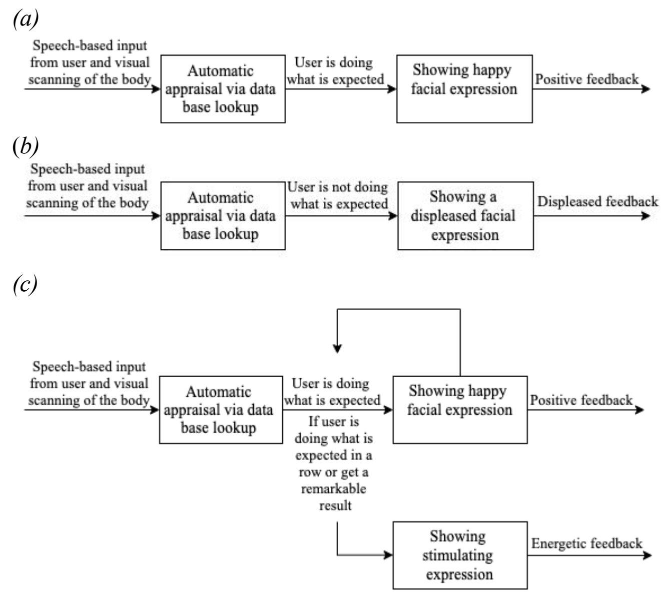

Figure 2. Illustrates MARC modules: *(a)* user is doing what is supposed, showing emotion of happiness; *(b)* user is not doing what is supposed, showing emotion of displeased; *(c)* user is doing a remarkable job, showing emotion of happiness and positive voice feedback.

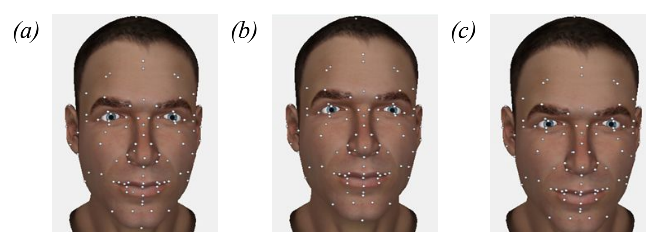

Figure 3. Facial expressions to display happiness.

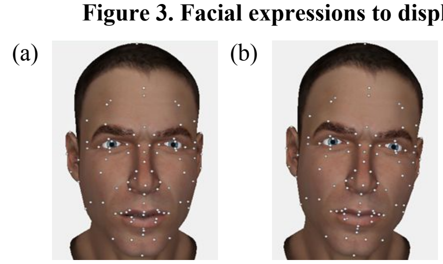

Figure 4. Facial expressions to display displeased.

The emotion of happiness is achieved by the employment of the cheek raiser and the lip corner puller action units, *see figure 3 (a)*. In addition, the head up, *see figure 3 (b)*, and head down action units, *see figure 3 (c)*, are applied to display approval.

The displeased emotion is achieved by applying the brow lowerer action unit, *see figure 4 (a)*, alongside with a head tilt left to illustrate discomfort, *see figure 4 (b)*.

DESIGN CONSIDERATIONS

Mockups

As we state earlier, the system is constantly recognizing the image received. When there is no movement during a period of time, the system will suspend itself and the image will become darker. The waiting time before suspension can be set by the user.

The moment when the user is recognized by the system, a dotted line will border his shape. The blue dotted line displayed in *figure 5* represents a neutral state when the user is not doing any movement related to the training.

When the user performs a movement related to the training plan, the system will analyze the position and search in the database for indicators. If the user's position matches the indicators in the database, a positive feedback will be produced. In *figure 6* it can be observed the green dotted line and a green shadow behind the virtual coach. This feedback will be accompanied by voice output.

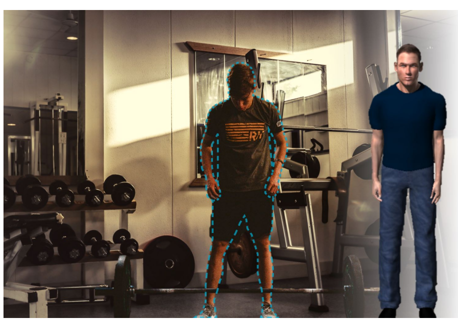

Figure 5. View of the system in a neutral state.

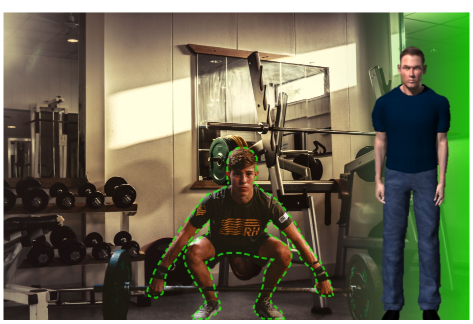

Figure 6. View of the system in a positive state.

If the user is executing a movement that does not match the indicators in the database, the dotted line and the shadow behind the virtual coach will become orange, as represented in *figure 7*. This feature aims to obtain the user's attention and to persuade him into making corrections. There will be voice feedback to specify the problem and to give instructions and advice on how the user can perform properly and obtain the right position.

Figure 7. View of the system in a displeased state.

IMPLEMENTATION WITH A MARC STATE MACHINE

The state machine of this system was implemented with a group structure. Each group does a set of specific tasks: "Greetings", "Stretching Part", "Rejection", "Emotional Response", "Exercises Part" and "Goodbye".

The "Greetings" is represented with three states. The first state is "Welcome". Its task is to say welcome to the user and ask him to start today's session. If the user says "Yes" then the coach goes to "Ready to start" state and finally goes to the "Stretching Part" group. If the user says "No" the system goes to the "Why not" state and then goes to the "Rejection Group", *see figure 8*.

The "Rejection" group handles the situation when the user says "No" to the system. The coach is going to persuade the user to start the today's session, *see figure 8*. When the user finally says "Yes", the coach will go to the "Stretching Part" group.

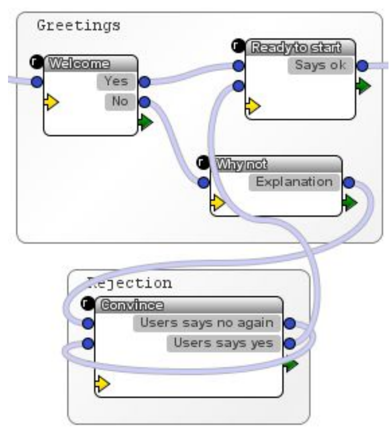

Figure 8. "Greetings" and "Rejection" groups.

The first state of the "Stretching Part" group is "Choose Stretch". It selects randomly a stretching exercise from a list that contains all the stretches. Then, the exercise is communicated to the user. Afterwards, the user must execute the exercise and select the key "*g*" if it was performed correctly, or select the key "*b*" if the exercise was performed inadequately, *see figure 9*. Subsequently, the system goes to the "Emotional Response" group to provide feedback to the user according to the performance. The process will occur five times before going to the next group of states, the "Exercise Part". This next group is essentially equivalent to the "Stretching Part" group. The difference is that the list used to randomly choose the exercises will now contain training exercises rather than stretches.

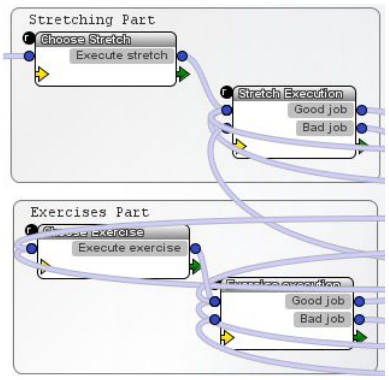

Figure 9. "Stretching part" and "Exercise Part" groups.

The "Emotional Response" group is the core of the system, as it provides an empathic response to the user. It is composed of three states, one for positive feedback, one for displeased feedback and another one for an energetic feedback (thrilled response). This last response is generated when the user executes correctly two consecutive exercises. The positive feedback is given when the user executes correctly the exercise or stretch. If the exercise or stretch is executed inadequately, then a negative feedback is generated, *see figure 10*. The process will occur ten times before going to the next group of states, the "Goodbye".

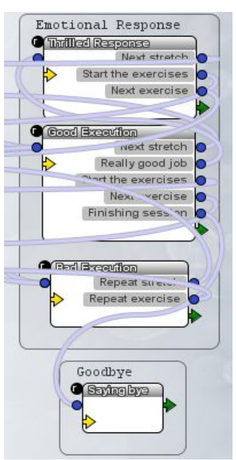

Figure 10. "Emotional Response" and "Goodbye" groups.

The "Goodbye" group contains only one state to provide a message to finish the session and then the system turns itself off, *see figure 10*.

RESULTS

The system initiates with a greeting from Sam, the virtual coach. The user is answering the virtual coach with voice-recognition with help of different keywords, as explained above. In the implemented system the user must interact with the keyboard to indicate how an exercise was executed.

User test

A group of 10 participants performed tests to explore the system. We provided them with standard sentences that the system could recognize. The main feedback obtained from the users was about the unsatisfactory accuracy of the voice recognition and the insufficient amount of recognizable sentences. The users appreciated the facial expressions displayed by the virtual coach and experienced the uncanny valley effect.

DISCUSSION

As a result of our team limited experience in the training context studied in this article, the implemented system works exclusively for the illustrative scenario explained previously. However, the state machine can be extended in the future to include more scenarios.

Furthermore, the essential difference with the illustrative scenario is the absence of image recognition to identify the users movements. To conduct successful tests with the system without employing said feature, we need to apply the Wizard of Oz testing method. This is a method used in prototype systems that uses a human to mimic functions that are not developed at the time of the testing [9].

When the user does the exercise, there will be a human person, referred to as the coach, observing the movement and judging its accuracy. Afterwards, the coach will give to the system text input so that the virtual coach can give the appropriate response and feedback.

Another aspect to consider is the level of satisfaction that the user may have with the system. The METUX model can be applied to evaluate if the needs of the user are being satisfied by the system [10]. For the first five spheres of experience (Adoption, Interface, Task, Behavior and Life) we can use the questionnaire templates [10] and deliver them in different moments for the user to give us feedback. We can measure Adoption during the first use of the system and Interface after the first week of use. The training activities of our system are not static, and can change as the user continues working with the system. For that reason we should assess Task a few sessions after a new exercise (task) is added to the routine. According to the results of this questionnaire, the system should adapt in case the user's needs are not being satisfied. Finally, Behaviour and Life can be evaluated after a few months of use, to ensure the user has enough experience with the system to observe real changes. As for the last sphere of Society, it is important to evaluate the effects the system can have on the work of real coaches in the weight lifting field.

The architecture presented in this article is partially similar to the SAIBA framework [11]. The recognition module used in our system differs from the Intent Planning module [11], as we do not analyze the user's input nor decompose it in characteristics. The voice or text input received by our system goes directly to the state machine, which performs as the Behavior Planning module[11]. However, the MARC tool allows the virtual coach to execute the selected behavior specified BML, as suggested by the SAIBA framework for the Behavior Realization module[11].

As we stated, there are features that were not implemented because of the limitation of time and of the software used to develop the virtual coach. The first feature is the movement analysis, that can be developed using a camera to read the user's exercise execution. The program could record the movement and analyse it through **the movement analysis module.** Inside this module the movements done by the user can be compared against the movements of a local record of the exercise. Therefore, our system can produce a response depending if the exercise was done correctly or not. The second feature is the recognition of emotions through the user's voice which will give us a better approximation to the SAIBA framework. This can be done using some natural language processor software, for example, **Tensorflow**. This tool can analyse a voice and recognize key words to assign a meaning to an user's response. Using tensorflow alongside the OCC module could allow the system to infer an emotion and provide a proper response.

The virtual coach presented can be evaluated with a group of weight lifting professionals. Their task would be to do a complete session with the coach and provide the following information: sentiments with their session with the coach, improvements compared to a session with a real coach.

AUTHORS

Angelica Acosta Arteta is an exchange student from Simon Bolivar University and is currently taking her Master degree at Télécom SudParis in Computer Science. Daniel Rodríguez is as well an exchange student from Simon Bolivar University and is also taking his Master degree at Télécom SudParis in Computer Science. Evelyn Bankell is an Erasmus exchange student from Linköping University and is currently studying at Télécom SudParis.

REFERENCES

[1] Kettunen, E., Critchley, W., and Kari, T. 2019. Can Digital Coaching Boost Your Performance? : A Qualitative Study among Physically Active People. In Proceedings of the 52nd Hawaii International Conference on System Sciences (HICSS 2019) (pp. 1331-1340). University of Hawai'i at Manoa. Retrieved from http://hdl.handle.net/10125/59574

[2] Courgeon, M., Clavel, C., Martin, J. 2014. Modeling Facial Signs of Appraisal During Interaction; Impact on Users' Perception and Behavior. International conference on Autonomous agents and multi-agent systems. Paris, France. Retrieved from https://hal.archives-ouvertes.fr/hal-01144001

[3] Petsani, D., Kostantinidis, E. and Bamidis, P. 2019. Designing an E-coaching System for Older People to Increase Adherence to Exergame-based Physical Activity. In: *Proceedings of the 4th International Conference on Information and Communication Technologies for Ageing Well and e-Health* (ICT4AWE 2018), 258-263. https://www.scitepress.org/Papers/2018/68215/68215.pdf

[4] Russell, J. A. 2003. "Core affect and the psychological construction of emotion." In: Psychological review 110.1, p. 145 from https://bing.bc.edu/james-russell/publications/psyc-rev2003.pdf

[5] Steunebrink. B, Dastani. M, Meyer. J. 2009. The OCC model revisited. Retrieved from http://www.idsia.ch/~steunebrink/Publications/KI09_OCC_revisited.pdf

[6] Tironi, A., Mainetti, R., Pezzera, N. M., and Borghese, A. 2019. An Emphatic Virtual Caregiver for assistance in exer-game-based rehabilitation therapies. Retrieved from https://ieeexplore.ieee.org/abstract/document/8882477

[7] van Kleef G. A., Cheshin A., Koning L. F., Wolfa S. A. 2019. Emotional games: How coaches' emotional expressions shape players' emotions, inferences, and team performance. In: *Psychology of Sport and Exercise* (Volume 41). 1-11. https://doi.org/10.1016/j.psychsport.2018.11.004

[8] WHO. 2010. *Global Recommendations on Physical Activity for Health*. Geneva: World Health Organization.

[9] Schlögl, S., Doherty, G., Luz, S. 2013. Managing Consistency in Wizard of Oz Studies:A Challenge of Prototyping Natural Language Interactions. Munich, Germany. Retrieved from https://www.researchgate.net/publication/259976870_Managing_Consistency_in_Wizard_of_Oz_Studies_A_Challenge_of_Prototyping_Natural_Language_Interactions

[10] Peters, D., Calvo, R., Ryan, R. 2018. Designing for Motivation, Engagement and Wellbeing in Digital Experience. In: *Frontiers of Psychology* (Volume 9). https://doi.org/10.3389/fpsyg.2018.00797

[11] Kopp, S., Krenn. B., Marsella. S., et al. 2006. Towards a Common Framework for Multimodal Generation: The Behavior Markup Language. Retrieved from: https://www.researchgate.net/publication/221588332_Towards_a_Common_Framework_for_Multimodal_Generation_The_Behavior_Markup_Language