Open Science Journal of Electrical and Electronic Engineering
2019; 6(1): 1-4
http://www.openscienceonline.com/journal/j3e

SMILE: A Verbal and Graphical User Interface Tool for Speech-Control of Soccer Robots in Ghana

Patrick Fiati

Electrical/Electronic Engineering, Cape Coast Technical University, Cape Coast, Ghana

Email address

patrick.fiati@cctu.edu.gh

To cite this article

Patrick Fiati. SMILE: A Verbal and Graphical User Interface Tool for Speech-Control of Soccer Robots in Ghana. *Open Science Journal of Electrical and Electronic Engineering*. Vol. 6, No. 1, 2019, pp. 1-4.

Received: February 8, 2019; **Accepted**: March 21, 2019; **Published**: April 10, 2019

Abstract

SMILE (Smartphone Intuitive Likeness and Engagement) application, a portable Android application that allows a human to control a robot using speech input. SMILE is a novel open source and platform independent tool that will contribute to the robot soccer research by allowing robot handlers to verbally command robots. The application resides on a smartphone embedded in the face of a humanoid robot, using a speech recognition engine to analyze user speech input while using facial expressions and speech generation to express comprehension feedback to the user. With the introduction of intuitive human-robot interaction into the arena of robot soccer, we discuss a couple specific scenarios in which SMILE could improve both the pace of the game and autonomous appearance of the robots. The ability of humans to communicate verbally is essential for any cooperative task, especially fast-paced sports. In the game of soccer, players must speak with coaches, referees, and other players on either team. Therefore, if humanoids are expected to compete on the same playing field as elite soccer players in the near future, then we must expect them to be treated like humans, which include the ability to listen and converse. SMILE (Smartphone Intuitive Likeness and Engagement) is the first platform independent smartphone based tool to equip robots with these capabilities. Currently, humanoid soccer research is heavily focused on walking dynamics, computer vision, and intelligent systems; however human-robot interaction (HRI) is overlooked. We delved into this area of robot soccer by implementing SMILE, an Android application that sends data packets to the robot’s onboard computer upon verbal interaction with a user.

Keywords

SMILE, HRI, ROBOTS, HUMANOID, ANDROID

1. Introduction

Robot Localization Module

It is well known that the estimation of the robot position from odometer is subjected to large errors in the long term. Consequently, additional sources of information must be used to determine the robot's position. In our case, we advocate for the use of vision. The appearance-based localization approach [1] is appealing for its simplicity. In this approach, the robot position is determined by a direct comparison between the image observed at a given moment and a set of images observed before. Note that, since images are directly compared to each other, we avoid the errors associated with the landmark detection procedures. The appearance-based localization departs from a training set including images shot at known positions. The dimensionality of these images is reduced using a standard PCA technique, producing an appearance-based map of the environment. When the robot is moving, the collected images are compressed using PCA and the resulting set of feature detectors is compared with those stored in the appearance-based map. The position of the map points closer to the current observation is used within a Markov process [2] to improve the estimation of the position of the robot. We have obtained good results combing this approach with the use of a particle filter estimate the probability distribution on the robot's position [3]. In our previous work, we used an omnidirectional camera. The problem of using this type of camera is that images are very sensitive to any change in the environment. Thus, to increase the robustness, within the \Ambience" project, we plan to use two cameras mounted on the head of the robot. The head can be moved to point the cameras in different directions. Therefore, in the presence of confusing images (images that do not match with any of the images in the appearance-based map) we can readily obtain a new image just moving this head. The idea is to select the head movement that is likely to provide the less ambiguous view. In this way, we will minimize the number of head movements necessaries to determine the robot's position [4]. An additional advantage of using a stereovision sys-tem is that it can provide depth maps that are less sensitive to change in illumination than usual intensity images. The objective of the navigation module is to drive the robot to the desired positions (expressed as abstract placements such as kitchen, bedroom, etc) avoiding collision with obstacles. There are two internal modules: the planner and the behavioral execution module. The planner generates sub goals from the robot's current location to the global goal position using a map. In previous work, a computationally efficient planner was developed based on the Wave Front Distance Field (WFDF) algorithm [5]. The straight path between two consecutive sub goals is obstacle free. This avoids that sub goals are too close together, which is not desirable in behavioral execution. The information provided by the localization module is used deter-mine the position of the robot and, thus, to keep track of the change in position already achieved [6]. The second component of the navigation architecture, the behavioral execution module, receives as in-put (from the planer) the desired relative displacement for the robot and determines the linear and angular speeds necessaries to perform it. Then, these speed scan be readily transformed to wheel motor commands [7]. The behavioral execution is implemented using a behavior-based control approach. Obstacles which are not in the map, both static and dynamic, possibly show up in front of the robot while moving. To avoid bumping into them, an avoid-obstacle algorithm is implemented. Ultrasonic sensors are used to detect these obstacles. The cooperation of the fast planner module and the behavioral execution one leads the robot to his goals [8].

2. Methods

As Norman pointed out in his 2001 keynote address on HRI, robots must have personality in order for humans to be able to predict and understand their behavior. In this way, humans cooperating with robots will be able to smoothly anticipate the robots' limitations and allow for them, in much the same way as people understand and work with each other's tendencies and limitations. One way to express personality, as well as task-relevant information, is through natural dialogue. While we are still a long way away from truly natural dialogue (as Norman says, the issue is not so much "speech recognition" as "language understanding"), by pairing emotional communication with natural language communication, we are making a step towards truly expressive robots. Such robots can work with humans in a way that non-technical users find comfortable and natural, allowing them to focus more on the task at hand than figuring out how to use the robot. Several service robots under development use verbal dialogue systems to receive and confirm commands, as well as to obtain assistance while navigating indoor environments. Such robots include Godot, Lino, and Cero. Godot in particular uses a sophisticated cognitive map and semantic analysis system to supplement its knowledge about its navigational environment from information gained during dialogue. Similarly, Lino uses emotional expressions as a part of its dialogue, and Cero uses simple gestures to indicate levels of understanding and intention. However, all of these robots use mechanical means of expressing emotion, which limits the complexity of their emotional indicators. Other robot developers have turned to screens and projected images for their robots' facial expressions, which are much easier and faster to animate. Baxter, for example, is a semi-humanoid industrial robot designed to work alongside humans in a factory setting, and its face is a screen that displays two emotional eyes. Human workers train Baxter by moving its compliant joints until it learns the task at hand; all the while, the robot's eyes reflect its internal state towards the task. Human "coworkers" of the Baxter robot view and treat it as a large child, indicating the effectiveness of its personality while being integrated into a cooperative environment. The SMILE app is the first attempt at using a Smartphone as a conversational and emotional interface with a task oriented robot. We hope to build on the work of others by exploring task-oriented dialogue and emotion/personality models, in order to enhance the behavior and soccer skills of our robots, as well as making these robots better suited for cooperative tasks involving human users.

3. Results and Discussion

3.1. Speech Recognition and Generation Methods

The application is separated into two simple layouts; one for the welcome screen and one for the emotional and conversational interface. Each layout is associated with a java class that implements an activity, which allows access to the user interfaces of the phone, such as the touch screen or microphone. The start activity is what appears when the app is launched [9]. When the start button is pressed, it will launch the face activity. Once the face activity has begun, the app is able to use Google Speech Recognition API to convert microphone input into text, which can be handled in several ways depending on the current SMILE listening mode. The SMILE listening modes described below are different routines for dealing with speech input [10].

1. Normal Listening Mode. Listen for keywords and phrases found in vocabulary. If input is found in vocabulary, use text-to-speech to output the corresponding response. If not, repeat the word back to the user.
2. Learning Listening Mode. Ask the user for a new word or phrase, ask the user for an emotion with which the learned item should be matched, and finally ask for a

response to output when the item is heard in the future.

3. Command Listening Mode. Listen for keywords and phrases found in the command mode vocabulary. If the word is a valid command, send the string over Wi-Fi network to the specified server port [11].

3.2. Command Listening Mode's Application to Robot Soccer

While SMILE has a few interesting listening modes, we are most concerned with command mode for use in robot soccer. In a game situation, command mode can be used to change parameters during game breaks without having to touch the robot. Practical applications for specialized command sets include changing strategy, walking technique, or any other robot behavior that could improve performance [12]. An important consideration when developing SMILE as a tool is that the user must assume that whatever he or she says will end up on the robot's onboard PC and the robot will act as intended [13]. Therefore, a robust algorithm must be designed for the specific command set. This command set will be unique for each team as there are differences between each team's spoken language, computing language, and intended robot behavior. Depending on the complexity of the command set, SMILE may be used to do initial filtering on the command before it reaches the onboard PC [14].

3.3. SMILE-to-Onboard PC Networking

When the user issues a command to SMILE, the transmission of a datagram packet containing that command is facilitated using UDP/IP, a transaction oriented protocol. This type of protocol is ideal for the simple messages passed from the smartphone client socket to the robot's onboard PC server socket [15].

4. Conclusion

In this paper, we discussed the SMILE app, a platform independent open-source tool, and its novel contribution to HRI for robot soccer purposes. By incorporating SMILE in the head of a soccer robot, humans are able to verbally communicate with the robot in an intuitive way that bypasses traditional forms of user input. Foremost, SMILE is able to network with the robot's onboard PC, allowing it to manipulate the robot's behavior. While SMILE is a new tool in the robot soccer arena, there are dynamics challenges associated with adding mass to a robot's head, which must be accounted for on a robot-to-robot basis. Nevertheless, with the excellent work being done in the area of walking dynamics, there is high potential for implementing a walking engine that sufficiently compensates for large head mass. Lastly, the foundation of the SMILE Android client and its accompanying C++ server use completely generic messages, making them tools that can be applied to countless practical applications within robotics. We released the SMILE Android app and the C++ server in the hopes that other researchers will follow in our footsteps to merge the fields of HRI with robot soccer in Ghana.

References

[1] John Paul Titlow, "For Robots Like Baxter, The Interface Becomes A Personality." Fast Company Labs. Retrieved September 29, 2014 [Online]. Available: http://www.fastcolabs.com/3009374/ for-robots-like-baxter-the-interface-becomes-a\ -personality

[2] Google, Inc. "Package Summary: android. speech," Android Developers, [Online]. Available: http://developer.android. com/reference/android/speech/package-summary. html (Accessed October 12, 2013.)

[3] Google, Inc. "Package Summary: android. speech. tts," Android Developers, [Online]. Available: http://developer. android.com/reference/android/speech/tts/ package-summary.html (Accessed October 12, 2013.)

[4] J. Cassell, \Embodied Conversational Agents: Representation and Intelligence in User Interface", AI magazine.

[5] A. J. Davison, M. Montemerlo, J. Pineau, N. Roy, S. Thrunand V. Verma,\Experiences with a Mobile Robotic Guide for the Elderly," in Proceedings of the AAAI National Conference on Artificial Intelligence 2002.

[6] P. Elinas, J. Hoey, D. Lahey, J. D. Montgomery, D. Murray, S. S. James and J. Little,\Waiting with Jos∂e, a vision-based mobile robot" in Proceedings of the 2002 IEEE International Confer-ence on Robotics and Automation Washington, DC, May 2002, pp. 3698-3705, 2002. 6

[7] ITEA Ambience project, http://www.extra.re-search.philips.com/euprojects/ambience/.

[8] B. J. A. Krƒose, N. Vlassis, R. Bunschoten and Y. Motomura, \Aprobabilisticmodelforappearance-basedrobot localization". Image and Vision Computing, 19 (6): 381-391, April 2001.

[9] B. J. A. Krƒose, N. Vlassis and R. Bunschoten, \Omnidirectional visionfor Appearance-based Robot Localization", in Lecture Notes in Computer Science, pages 39-50. Springer, 2002.

[10] Marcel Missura and Sven Behnke. "Self-stable Omnidirectional Walking with Compliant Joints." In Proceedings of 8th Workshop on Humanoid Soccer Robots 13th IEEE-RAS International Conference on Humanoid Robots (Humanoids), Atlanta, GA, 2013.

[11] J. M. Porta, B. Terwijn and B. Krƒose, \efficient Entropy-Based Action Selection for Appearance-Based Robot Localization". In Proc. IEEE Int. Conf. on Robotics and Automation, Taipei, May 2003.

[12] Hafez Faraz, Mojtaba Hosseini, Vahid Mohammadi, Farhad Jafari, Donya Rahmati, Dr. Esfandiar Bamdad, "Baset Teen Size 2014 Team Description Paper." (2014). Available: http://application.germanteam.org/upload/7708a96f5dbd4e541 e64bacc933c3db21f90410c/Baset-Teen_TDP.pdf

[13] D. A. Norman, "How might humans interact with robot's Human Robot Interaction and the Laws Of Robotology, Notes for a Keynote Address to the DARPA/NSF Workshop on Human? Robot Interaction, San Luis Obispo, CA. (2001, September). [Online]. Available: http://www. jnd.org/dn.mss/Humans_and_Robots.html (July 2002).

[14] Christian Theobalt, et al. “Talking to Godot: Dialogue with a mobile robot.” Intelligent Robots and Systems, 2002. IEEE/RSJ International Conference on. Vol. 2. IEEE, 2002.

[15] Ben JA Krse, et al. “Lino, the user-interface robot.” Ambient Intelligence. Springer Berlin Heidelberg, 2003.