In Proceedings of 9th Workshop on Humanoid Soccer Robots
14th IEEE-RAS International Conference on Humanoid Robots (Humanoids), Madrid, Spain, 2014.

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

Adam B. Stroud , Elise Russell, Raoul Chinang, Kellen Carey,
John C. Williams, Darryl Ramgoolam, and Andrew B. Williams, PhD
Humanoid Engineering and Intelligent Robotics (HEIR) Lab
Marquette University Opus College of Engineering
Milwaukee, Wisconsin USA
adam.stroud@marquette.edu, andrew.williams@marquette.edu
http://www.eng.mu.edu/abwilliams/heirlab/

***Abstract*— In this paper, we discuss 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. Running on a commercially available smartphone, the SMILE application was first implemented on the MU-L8 TeenSize platform, as shown in this video `http://youtu.be/EQfZ_BoDF5c`.**

***Index Terms*– Human-Robot Interaction (HRI), SMILE (SMartphone Intuitive Likeness and Engagement), TeenSize (80-130 cm tall humanoid)**

I. INTRODUCTION

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 includes the ability to listen and converse. SMILE (Smartphone Intuitive Likeness and Engagement) is the first platform independent smartphone-based tool to equip robot's 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 on-board computer upon verbal interaction with a user. SMILE was developed to be a quick way to evoke robot behavior through verbal commands with wide ranging applications for all robotics, especially the fast-paced environment of RoboCup. For example, in today's RoboCup competitions when a robot handler needs to alter a player's soccer behavior, they can use buttons on the robot's exterior for quick fixes such as state control, but more specific adjustments to system parameters must be made using a remote connection to the robot's onboard computer. This can be quite time consuming, and detracts from the robot's sense of autonomy. Hence, SMILE is an auxiliary application on a smartphone embedded in the face, which can allow teams to bypass telepresence methods with specialized verbal commands.

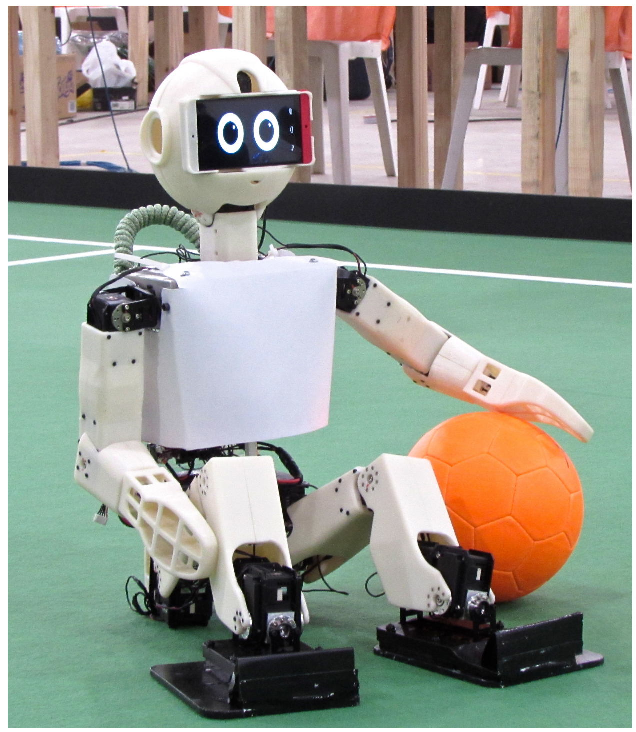


Fig. 1: The SMILE application implemented in Sonny, a version of the MU-L8 TeenSize platform.

In the following section, we discuss related work in the field of HRI as it pertains to task-oriented environments. We then discuss SMILE's integration into a humanoid soccer platform in two main sections: software architecture of SMILE and considerations on the physical incorporation of a smartphone in the head of a robot. Finally, we present a preview of practical improvements to SMILE that could further incorporate HRI into soccer robots.

II. Related Work

As Norman pointed out in his 2001 keynote address on HRI [1], 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 [2], Lino [3], and Cero [4] [5]. 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 [6].

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.

III. SMILE Application Architecture

SMILE [15] is built upon Google's open-source, Linux-based Android mobile operating system, which provides a framework for SMILE's key functions, including speech recognition, speech generation, and wireless networking. Although SMILE's foreground uses blinking eye animations that respond to certain keywords and changes in orientation, the focus of this section will be on the background process that deals with user input.

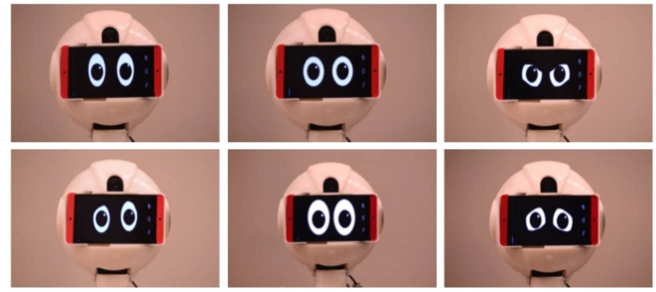


Fig. 2: Several appearances contained within SMILE's face activity layout can convey information about the system to the user. Additionally, the developer can choose how SMILE visually responds to certain stimuli.

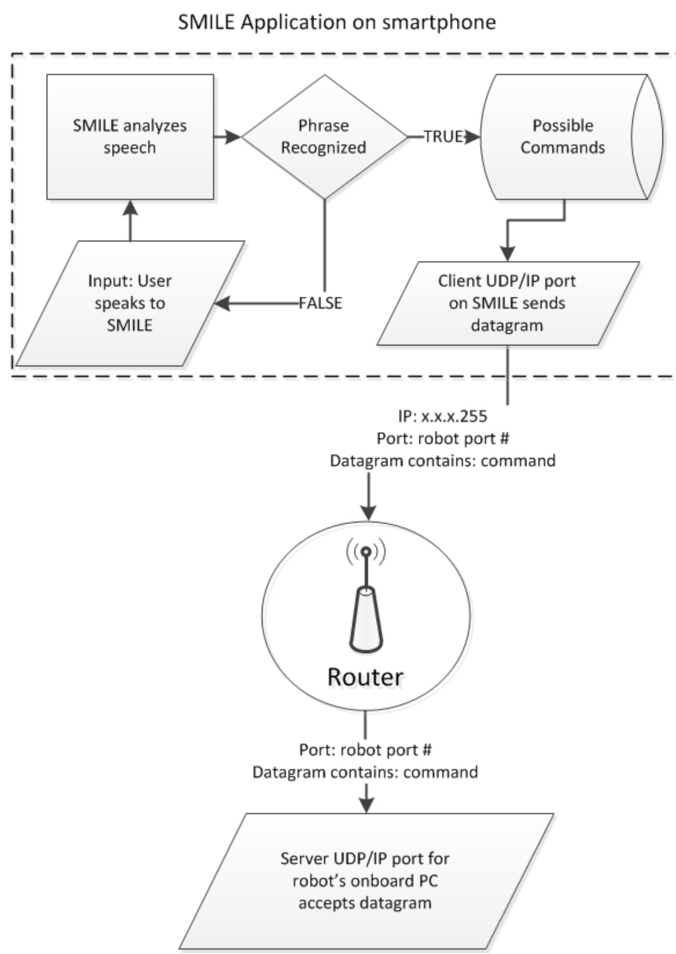


Fig. 3: The transmission of a single datagram occurs when the user issues a command to SMILE. The client socket sends the datagram with a broadcast IP address and server port specified by the onboard PC.

A. 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 shown in Figure 2. 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. 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 [7] 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.

- *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 [8]. If not, repeat the word back to the user.
- *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.
- *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.

B. 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.

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. 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. In Figure 4, we demonstrate a set of possible commands and their format as they arrive on the onboard PC.

C. 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 [9]. This type of protocol is ideal for the simple messages passed from the smartphone client socket to the robot's onboard PC server socket. The client-server relationship between SMILE and the onboard PC shown in Figure 3 is one-way, that

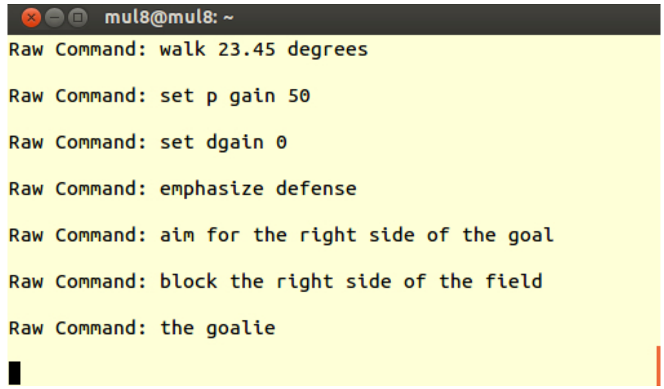


Fig. 4: SMILE has accurate speech recognition that allows the user to reliably send entire clauses and even numbers with up to 2 decimal places. In this example, the onboard PC's server socket, written in C++, is receiving unfiltered datagrams from SMILE.

TABLE I: Effects of a Smartphone on Mass Distribution of MU-L8

Metric	Without Smartphone	With Smartphone
Total Mass	7400 g	7539 g
Center of Mass	444 mm	454 mm
Total Height	908 mm	-

is datagram transmission flows from the smartphone client, through a router, and finally to the onboard PC's server. A two-way connection between the devices, not featured in the current version of SMILE, would simply require the implementation of both a server and a client on each device. This would be useful if SMILE were applied as a quick way to display statistics about the robot's internal state, such as battery level or motor temperatures.

IV. Physical Implementation of SMILE in Bipedal Humanoids

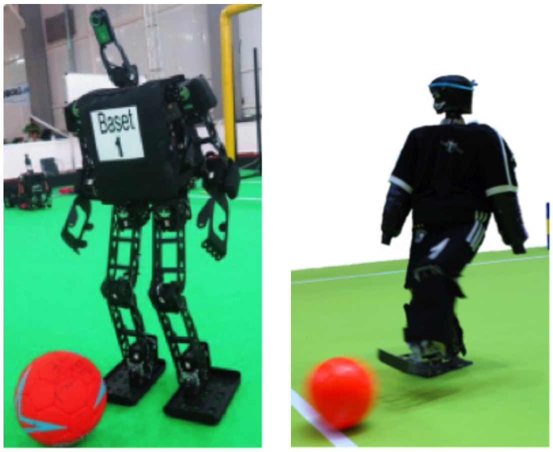
As we define the term humanoid robot, let us not forget that the head and eyes play an important role in anthropomorphizing a robot [10].

A. Protective Placement of Embedded Smartphone

In the MU-L8 humanoid, the smartphone is protected by a detachable 3D-printed pocket, which is nestled into the head's anterior. In the event that the robot falls directly on its face, the ABS plastic head can absorb considerable damage while protecting the chassis and screen of the smartphone. As evidence that the smartphone was durable enough for competition, the detachable pocket tended to break apart after several forward falls, yet the smartphone remained unharmed.

B. Dynamics Challenges of Added Head Mass

While SMILE and its smartphone host introduce exciting new capabilities to the humanoid, the mass of the phone also heavily affects the dynamics of the robot. Table I compares the smartphone's effect on MU-L8's center of mass and total mass. The 139 gram LG Nexus 4 smartphone and 499



(a) Baset TeenSize (b) Dynaped [14]

Fig. 5: Both (a) and (b) are examples of soccer robots that use head designs that have minimal impact on walking dynamics.

gram head comprise 8.4% of MU-L8's total mass of 7539 grams, almost anthropomorphically identical to the average human head, which accounts for 8% total body mass [11] . Consequently, with a higher center of mass, the robot will be less balanced and the robustness of the system's dynamic walking engine will be challenged when the robot walks and pivots its head simultaneously.

Such a challenge is appropriate for robots that exhibit fast and highly controlled walking patterns such as Baset TeenSize [12] and NimbRo's Dynaped [13], both shown in Figure 5. These are examples of robots that minimize top-heaviness and unpredictable head movements by using low-mass and fixed-pitch head designs, respectively.

V. Future Work

Ideally, all RoboCup robots will have intuitive HRI abilities such as the ones presented in this paper. While this is an ambitious goal, there is no technological barrier preventing SMILE app from being more powerful or accessible. Hence, we propose three future projects that would solve the challenges associated with the current version of SMILE.

- *Harnessing smartphone IMU to improve walking.* Predicting the head's effect on balance during walking is a major challenge associated with a heavier head. Given the 6-axis IMU in many smartphones, the SMILE app can be used to predict the head's position and help the walking engine compensate for imbalances caused by quick jerks of the head during ball tracking.
- *Bidirectional communication between smartphone and onboard PC.* Another current weakness of SMILE is its reliance on a stable wireless network. Hence, the foremost task for SMILE is to switch from UDP/IP to a Bluetooth or USB serial protocol, thus introducing a bidirectional flow of information. Given the ability to communicate back and forth between SMILE and the onboard PC, there are several ways to present information about the system. For example, SMILE would be able to give visual or verbal status reports on battery level, motor temperature, or any other internal statistic. The purpose of developing abilities such as these is to make natural interaction with humans purposeful and efficient in the context of robot soccer.
- *Customizable appearance of SMILE.* In the near future, the user will be able to choose an appearance profile from a selection of eyes varying in color, style, and size. While this is a purely aesthetic feature, it is important that humans can easily distinguish one robot from another. This future development was motivated by Sonny and Forest, who are identical models of the MU-L8 platform.

VI. Conclusions

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, as seen with the MU-L8 TeenSize platform.

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.

Acknowledgment

This project was supported by the John P. Raynor, S.J, Endowment, Opus College of Engineering, and several generous donations from private donors and industry sponsors.

References

[1] 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).

[2] 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.

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

[4] Anders Green and Kerstin Severinson-Eklundh. "Task-oriented Dialogue for CERO: a User-centered Approach." Robot and Human Interactive Communication, 2001. Proceedings. 10th IEEE International Workshop on. IEEE, 2001.

[5] Kerstin Severinson-Eklundh, Anders Green, and Helge Httenrauch. "Social and collaborative aspects of interaction with a service robot." Robotics and Autonomous systems 42.3 (2003): 223-234.

[6] 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
[7] Google, Inc. "Package Summary: android.speech," Android Developers, [Online]. Available: http://developer.android.com/reference/android/speech/package-summary.html (Accessed October 12, 2013.)
[8] 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.)
[9] J. Postel, "User Datagram Protocol," RFC 768, USC/Information Sciences Institute, August 18, 1980.
[10] Cynthia L. Breazeal, "Designing sociable robots with CDROM." MIT press, 2004.
[11] Danny Yee. "Average Human Head Weight." Department of Anatomy and Histology, University of Sydney. June 2006. [Online]. Available: http://danny.oz.au/anthropology/notes/human-head-weight.html
[12] Hafez Farazi1, Mojtaba Hosseini2, Vahid Mohammadi,Farhad Jafari, Donya Rahmati, Dr. Esfandiar Bamdad, "Baset Teen-Size 2014 Team Description Paper." (2014). Available: http://application.germanteam.org/upload/7708a96f5dbd4e541e64bacc933c3db21f90410c/Baset-Teen_TDP.pdf
[13] 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.
[14] Sven Behnke, Marcel Missura, Hannes Schulz. "NimbRo TeenSize 2012 Team Description (2012). [Online]. Available:" http://www.ais.uni-bonn.de/nimbro/Humanoid/papers/TDP_NimbRo_TeenSize_2012.pdf
[15] Elise Russell, Adam Stroud, Darryl Ramgoolam, Jerrell Jones (2014, October). SMILE Application and C++ Server. [Online]. Available: https://code.google.com/p/smile-application-mul8/