Prezence - Kinetic Feedback Module

Daryl Ma [dm2913]

***Abstract*—Using Human-Robotic Interaction as a means of providing presentation feedback is an approach that is yet to be done in significant detail. In this report, an investigation is carried out into previous approaches for providing presentation feedback to a presenter, as well as preceding techniques for non-verbal robotic communication through Kinesics. Drawing on these results, a novel approach combining presentation feedback with robotic kinesics is designed with the NAO robot. This will be implemented as part of the modules in the Prezence.**

I. Introduction

Communicating and presenting well is a valuable tool to have in order to accurately convey important ideas. With enough practice, a person's communication fluency and presentation skills could be improved and refined. Therein lies the value of the Prezence[14], a robotic presentation trainer capable of providing real-time responses as a means of feedback to a user, thus allowing a user to improve upon their presentation skills.

A user would have his presentation skills evaluated using metrics devised in the design report[14], data which the **Central Processing Module** would receive from the **Computer Vision** and **Speech Processing Modules**. The **Central Processing Module** then determines whether any action by the presenter requires immediate correction. This would result in the **Central Processing Module** triggering a feedback to be performed through the **Kinetic Feedback Module**.

Hence, the **Kinetic Feedback Module** is the organ of the Prezence dedicated to providing presentation feedback to the user. This will be undertaken through a robotic interface, and be used to test the hypotheses that this form of feedback would be useful in improving a user's presentation skills, provides an improvement upon other presentation feedback methods as well as boosting a user's confidence and comfort in their presentations.

II. Module Requirements and Challenges

This module would focus on the human-robotic interaction(HRI) between the user and Prezence. Of particular focus would be determining the optimal method of transmitting information as feedback between the Prezence and the user. Determining this would require an in-depth understanding into social robotics, as well as the psychology behind presentation feedback.

In order to overcome these challenges, this paper will look into previous methods of presentation feedback, whether real-time or post-speech; Designing the robotic gestures by looking at other similar projects; and finally, by combining the findings from both fields, design a robotic presentation feedback system. Ultimately, this is critical for the entire project, given that the aim would be to let the presenter respond and learn from feedback.

III. Background

Background research was done on two separate topics. The first topic consists of existing presentation feedback systems. Most of these systems involved a visual interface, as will be elaborated on in **Section IIIA**. The below-mentioned papers were useful for providing an idea into the factors involved for delivering effective presentation feedback. The second topic was looking into existing research regarding using a robot to communicate with humans. This has been done extensively, especially with the NAO. It was key for this experiment in order to decide the means of communication for the real-time feedback.

A. Presentation Feedback

1) Rhema: Google Glass Interface: Past designs on methods involving presentation feedback were looked into, from which two papers particularly stood out. The first was Rhema, a Google Glass application developed by the University of Rochester which provided a speech-based visual feedback[9].The aim of this presentation feedback method was to serve as a real-time public speaking coach that monitored the user's speech volume and pace, and pointing out the errors immediately through the visual display on the Google Glass. By comparing continuous streams of feedback against sparse feedback, their experiments showed that sparse feedback was preferred due to the large cognitive load shouldered by a presenter during a presentation.

One drawback of this method was that a wearable communications could distract someone speaking to the user[2]. That meant that having an additional tool for the user could serve to be more of a distraction for the audience rather than an aid to the presentation.

2) Video Panel Real-Time and Haptic Interface: The second method involved using a monitor display to mirror the Presenter's image and provide real-time feedback through a graphical and text interface using a video panel, as well as haptic feedback through a vibrating wristwatch, designed by academics from the Open University of Netherlands[4]. The feedback method involved using a mirrored image of the presenter displayed on a video screen, with the instruction transmitted through the vibrating wristwatch in addition to a text overlayed over the mirrored image.

In this experiment, the types of presentation feedback is separated into corrective and interruptive. Corrective feedback indicate errors that occurred once or twice, and hence are corrected with a light vibration of the wrist watch as well as a simple visual display. Interruptive feedback is done for errors that are repeated multiple times; The program is then paused, waiting for the user to correct the said error before it continues.

Their experiments were useful into looking for ways to escalate the level of seriousness of an error for the presenter. Given that most people are susceptible to habitual behaviours, escalating the feedback method could ensure that these users would take greater notice of their errors. For our case, this may involve additional functions that in the interest of time may not be achievable, but will be explored in future implementations.

Another point mentioned is a preferred delay between successive feedbacks to be around 6 seconds[4] in order to avoid distracting the user. This would be a useful gauge for the delays between successive Prezence feedback. Given that haptic feedback is not employed in the Prezence design, which relies upon immediate attention given the vibrating feedback, a sparser feedback method could have a shorter delay time. Finally, they also touched on the importance of having a robotic trainer that was interesting enough for the user to enjoy using as a presentation tool.

Both experiments noted that overwhelming distractions from presentation feedback caused unnatural and awkward behaviours. Hence, this has to be kept in mind while designing the feedback system. Another paper also showed the effectiveness of real-time feedback as opposed to post-speech feedback [1][5]. This meant that a balance had to be found such that real-time feedback given would be sparse yet effective, with the post-speech feedback consisting of the more complex information.

3) Presentation Feedback Guidelines: The following guidelines are devised from the research completed on presentation feedback methods:

- Real-Time feedback given does not overwhelm or distract the user
- Real-Time feedback should be on issues user can control with ease
- Real-Time feedback should have a delay time between successive feedbacks (Around 6 seconds)
- Large amounts of feedback should be delayed to the post-speech period

This guidelines will be followed in the high-level design of the **Feedback Module**.

B. Robotic Kinesic Communication

1) Choice of Robot: In order to assist with the comfort of the user, the humanoid NAO is chosen as the medium for the feedback. The NAO has the advantage of being able to perform realistic gestures signaling to the user and being as close to a humanoid shape as possible to represent a listening audience. The myriad of libraries available to the NAO such as the NAOaudio library and the NAOmotion also made it particularly favourable in terms of the amount of time required to utilise the robot.

Having a physical presense is also shown to be an advantage against a video-based feedback [4]. In the experiment by the department of social robotics in Yale University, participants were shown to be more aware of the physical presence of a robot as opposed to a video feed, and were also more likely to listen to requests made by a physical robot. This would hopefully allow the feedback to be delivered more adequately to the user.

2) Kinesic Communication: Kinesics involves the intepretation of the gestures and body language in order to communicate. Robotic communication through kinesics has been explored in multiple papers[8][10][6][7]. These experiments together showed that communication through robotic gestures was definitely possible, even through different mediums such as the NAO or a teddy bear robot[3]. Experiments done by Paul Bremner and Ute Leonards on robotic avatars showed that iconic gestures performed by robots were comprehensible even without speech accompaniment[8][10]. This is a vital factor, since gesturing would be key to the presentation feedback method in situations when speech is not used. It was also shown that by having participants evaluate gestures not on their meaning, but by what action is required as a response, made understanding the gesture much easier.

It is crucial to note the limitation of the NAO due to the lack of fully articulated hands[10]. Hence, gestures requiring hand shapes might be too ambiguous to be correctly identified, and has to be taken into account when designing gestures for the Prezence.

3) Robotic Kinesic Communication Guidelines: From these papers, we obtained the following guidelines for the Robotic Kinesic Feedback:

- Robotic Kinesic feedback has to be based on iconic or easily recognisable gestures.
- Robotic Kinesic feedback would be a gesture meant for the user to respond to.

IV. High-Level Design

The design of the **Kinetic Feedback Module** will involve the following decisions:

1) The real-time feedback mechanism
2) The post-speech feedback mechanism

In this section, the design decisions behind the mechanisms will be elaborated on and justified.

A. Real-Time Feedback Mechanism

As mentioned earlier, the feedback given to the user has to to be subtle enough to not distract the user, and also be clear and distinct enough for the user to correct themselves quickly enough.

1) Presentation Feedback Method: The real-time feedback works by the user noticing that the Prezence stops being in its normal standing position. The Prezence would then go to 1 of the 8 pre-defined gestures indicating a flaw in the presentation for 6 seconds. The user would then take note of the gesture during this period, and try to correct the presentation based on the gesture. The NAO would then return to its normal standing position. In this way, presentation feedback is provided to the user by having the user respond to an action done by the Prezence. This method is much more subtle than using a haptic feedback or by having a video feed for the user. Users could also choose whether to listen to or ignore the feedback, but the key point here is that they would not be distracted in order for them to persist with the presentation. Since there is post-speech feedback, any key information ignored during the real-time feedback could be communicated during the post-speech feedback mechanism.

Gesture List			
No.	Type	Audio and Visual Cues Detected	Gesture Reaction
1	Computer Vision	Lack of Eye Contact	Getting Attention(Single Hand Wave)
2	Computer Vision	Excessive Hand Movements	Standing at Attention(Hands Behind)
3	Computer Vision	Poor Body Posture	Leaning Over
4	Speech Processing	Clarity - Unclear	Hands Shrugging
5	Speech Processing	Pace - Too Fast	Arms Raised Up and Down
6	Speech Processing	Pace - Too Slow	Bored(Fist To Chin)
7	Speech Processing	Volume - Too Loud	Covers Ears
8	Speech Processing	Volume - Too Soft	Tilts Head and Cups Ear
9	NA	Listening State	Stands(Nods every 30s)

Table I. GESTURES FOR REAL-TIME FEEDBACK

2) Communication Method: The communication method decided upon was kinesics feedback done through postures by the Prezence. This is chosen over a speech-based real-time feedback in order to prevent the user from being distracted by the Prezence, as well as to accurately model a normal audience reaction when faced with similar situations. Although a speech-based feedback would be clearer and distinct, it is unlikely that a person would be interrupted in an actual presentation. Furthermore, a speech interjection during a presentation is distracting to the presenter. An additional advantage through the use of kinesics is that the user can be trained to take note of visual cues by observing the body postures and language in an audience.

3) Gesturing: The list of the 9 visual gestures can be shown in **Table 1** below.

The gestures chosen were based on what a person would feel given they were watching a presentation with the given actions. Hence, the Prezence would react, allowing for the user to heed the visual cue and thus correct himself. Note that these gestures are not finalised, and would be adapted based on the motion of the NAO as well as how the presenter would react towards these gestures.

B. Post-Speech Feedback Mechanism

The purpose of the post-speech feedback is to provide sufficient information on the performance of the user, such that the user can improve his performance for future presentations. This would entail a summary of the overall performance of the speaker as measured through the metrics stated in the design report given through a speech by the Prezence[14].

In this segment, the Prezence would behave like a tutor, rather than an audience. For metrics in which the presenter performed well, the Prezence would give a reply stating that the Presenter performed admirably, in order to encourage similar actions in the future. For metrics in which the Presenter performed poorly, the Prezence would motivate the Presenter to improve that feature.

Given that the metrics are only a way of evaluating the presentation, and may not be an accurate method of doing so, it is key that the Prezence post-speech feedback remains encouraging and supporting throughout. As mentioned in [4], the user should enjoy the Prezence as a tutor, and not be discouraged by it.

V. IMPLEMENTATION

The process behind the software required for programming the real-time feedback gestures as well as the post-speech feedback will be detailed in this section.

A. Software Environment

The software required for the **Kinetic Feedback Module** would be Choregraphe in order to accurately simulate the postures. The programming language used for the scripts would be Python. For intermodular communication, text files will be used from which the Python script will read from and write to. If time permits, the Robot Operating System(ROS)[15] will be utilised for its libraries and tools to optimise the communication process between the modules.

B. Real-Time Feedback

The flow diagram in **Figure 1** displays the mechanism for the real-time feedback. The Prezence would initially stand at attention, while continually checking for an input text file. If an input is detected, the data is decoded for the gesture meant to be performed, before going back to the standing state. This will be done through the use of a Python script, with each of the gestures consisting of a function that is called depending on the input file.

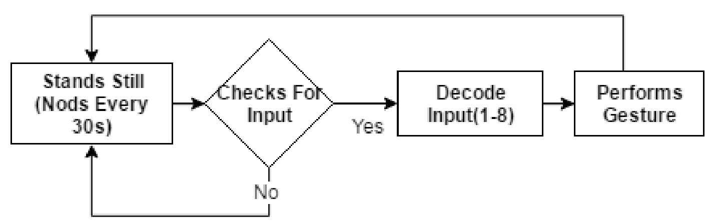

Figure 1. Flow Diagram for Real-Time Feedback Mechanism

C. Programming the Gestures

The NAO contains multiple modules that assist with the motion and the speech. Predefined postures can be found in the *ALRobotPosture()* module such as crouch, lying back, sit, stand, and stand zero. These postures could be utilized as gestures. However, in order to try to accurately mimic a real audience reaction, the gestures used would be of a unique design.

During the listening state of the Prezence, the *ALAutonomousMoves()* module would be utilised. This module contains a function that would simulate breathing motions, and ensures that the Prezence appears animated while listening to the Presenter.

The posture of the NAO can be controlled through 2 methods. The first method, Joint Control[13], uses joint angles in order to determine the motion. These joints consists of: *HeadYaw, HeadPitch, ElbowYaw, ElbowRoll, ShoulderPitch, ShoulderRoll* In order to effect the gestures, these joints can be controlled through the *motionProxy.setAngles()* function, where the joint angles are specified in the Python script. In order to specify these angles, it is paramount to know the degree of freedom of each joint as shown in **Figure 2**.

The second method, Cartesian Control[12], involves the use of effectors where the positions of a body part is defined relative to reference postion. The effectors consists of: *Head, LArm, RArm, LLeg*, among others. The desired position is input into this function *motionProxy.getTransform()*, from which a vector is returned that calculates the joint positions through the use of inverse kinematics solvers. This vector is then input into the function *motionProxy.transformInterpolations()* for the new arm positions to be moved.

However, calibration using these 2 methods required a long time, and both methods were eventually discarded for the 'puppeteering' method in the Animation Mode[11].By manually moving the posture of the NAO to the desired gesture, the required joint angles and motions are then monitored on Choregraphe and recorded in a keyframe. This speeds up the processing of slowly adjusting the angle using the Joint Control and Cartesian Control method to exact the correct body posture. This 'puppeteering' method is used to program all the gestures, and sped up much of the development time required for the **Kinetic Feedback Module**.

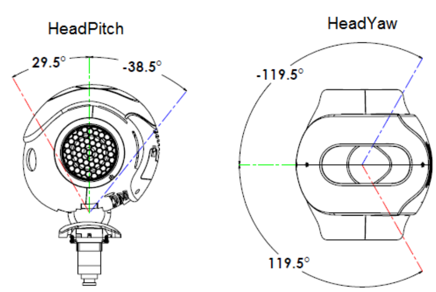

Figure 2. Head Pitch and Head Yaw Degree of Freedom

D. Post-Speech Feedback

For the initial implementation, these would simply involve the Prezence giving an overview of the presentation as measured by the metrics. This could be done effectively through the use of the *ALTextToSpeech* module in the NAO.

The overall statistics calculated and the performance indicator would be obtained in the text file sent from the Central Processing Module at the end of a presentation. These metrics would then be placed in a standard text file in the Python script, and read out by the NAO.

Some examples of the type of speech given is shown:

1) Good Range: "You looked at me `head gaze statistic` of the time. That was good! Keep on doing that next time!"
2) Below Average Range: "You spoke too loud `speech volume statistic` of the time. That wasn't so good, try to improve it for your next presentation!"
3) Average Range: "You spoke too quickly `speech pace statistic` of the time. That's alright, though there's room for improvement!"

These standard texts will be used to provide feedback on the metrics to the user after a presentation.

VI. Conclusion

This design report demonstrates a method for the design of a robot-based kinesic and speech feedback. In the event that kinesics proves to be an excellent method in communicating feedback to people, the use of this design in presentation-aided robots would be another step forward for a closer interaction between humans and robots.

References

[1] Paul E.King, Melissa J.Young, and Ralph R.Behnke. "Public speaking performance improvement as a function of information processing in immediate and delayed feedback interventions". In: *Communication Education* 49.4 (2000), pp. 365–374.

[2] Gerard McAtamney Caroline Parker. "An examination of the effects of a wearable display on informal face-to-face communication". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2006), pp. 45–54.

[3] Jamy Li. "Gesture Communication in Social Robots:A Case Study of a Robotic Teddy Bear". In: *Thesis, University of Toronto, Canada* (2008).

[4] A Bainbridge et al. "The Benefits of Interactions with Physically Present Robots over Video-Displayed Agents". In: *International Journal of Social Robotics* 3.1 (2011), pp. 41–52.

[5] Ligia Batrinca et al. "Cicero - Towards a Multimodal Virtual Audience Platform for Public Speaking Training". In: *International Workshop on Intelligent Virtual Agents* 8108 (2013), pp. 116–128.

[6] Chien-Ming Huang and Bilge Mutlu. "Modeling and Evaluating Narrative Gestures for Humanlike Robots". In: (2013).

[7] Allison Sauppé and Bilge Mutlu. "Robot Deictics: How Gesture and Context Shape Referential Communication". In: (2014), pp. 342–349.

[8] Paul Bremner and Ute Leonards. "Efficiency of Speech and Iconic Gesture Integration For Robotic and Human Communicators - a Direct Comparison". In: *Robotics and Automation (ICRA), 2015 IEEE International Conference on* (2015).

[9] M.Tanveer, Emy Lin, and Mohammed Hoque. "Rhema: A Real-Time In-Situ Intelligent Interface to Help People with Public Speaking". In: *Proceedings of the 20th International Conference on Intelligent User Interfaces* 1.1 (2015), pp. 286–295.

[10] Paul Bremner and Ute Leonards. "Iconic Gestures for Robot Avatars, Recognition and Integration with Speech". In: *Front Psychology* 7.183 (2016), pp. 1999–2006.

[11] Alderaban Documentation. *Animation Mode*. URL: http://doc.aldebaran.com/2-1/software/choregraphe/animation_mode.html#animation-mode.

[12] Alderaban Documentation. *Cartesian Control*. URL: http://doc.aldebaran.com/2-1/naoqi/motion/control-cartesian.html#control-cartesian.

[13] Alderaban Documentation. *Joint control API*. URL: http://doc.aldebaran.com/1-14/naoqi/motion/control-joint-api.html.

[14] Daryl Ma et al. *Prezence - Robotic Presentation Trainer*.

[15] ROS.org. *ROS Documentation*. URL: http://wiki.ros.org/.