Social Gaze Model for an Interactive Virtual Character

Bram van den Brink, Christyowidiasmoro and Zerrin Yumak ✉

Utrecht University, Princetonplein 5,
de Uithof, Netherlands
a.c.vandenbrink@uu.nl
c.christyowidiasmoro@uu.nl
z.yumak@uu.nl

Abstract. This paper describes a live demo of our autonomous social gaze model for an interactive virtual character situated in the real world. We are interested in estimating which user has an intention to interact, in other words which user is engaged with the virtual character. The model takes into account behavioral cues such as proximity, velocity, posture and sound, estimates an engagement score and drives the gaze behavior of the virtual character. Initially, we assign equal weights to these features. Using data collected in a real setting, we analyze which features have higher importance. We found that the model with weighted features correlates better with the ground-truth data.

Keywords: Gaze model, Engagement, Situated interaction

1 Introduction

Gaze movement is important for modeling realistic social interactions with virtual humans. While gaze animation based on low-level kinematics is well-studied, autonomous generation of gaze at the high-level during social interactions and in real settings still remains as a challenge [1]. One of the open problems is how to drive the gaze behavior of an interactive virtual character situated in a real environment, i.e. a virtual receptionist. It requires understanding which user has an intention to interact with the virtual character, in other words which user is more engaged. The users might be approaching the virtual character alone, in groups or they might just be passing by.

Recognition of goals, intentions and emotions of other people is important for a fluent communication. If one has to give human-like capabilities to artificial characters, they should also be able to predict the intentions of others. In this paper, we focus on the engagement detection problem as a prerequisite to initiating a conversation with a user and propose a model to autonomously drive the gaze behavior of the virtual character. Fig. 1 shows our Virtual Character Sara interacting with a group of users.

Fig. 1. Virtual Receptionist Sara interacting with users

Previous work modeled engagement based on heuristic rules [2][3]. It has been shown that machine learning approaches [4][5] outperform the basic heuristics. Both approaches have advantages and disadvantages. While the former does not involve extensive validations of their model, the latter depends on huge data collection and analysis efforts. An overview of multi-party interactions and a discussion on open research challenges can be found in our previous work [6]. In this research, our contribution is two-fold: (1) We present a practical and general engagement model combining multiple behavioral cues to drive the gaze of an interactive virtual character. (2) We find the importance of these behavioral cues based on data collected in a real environment.

We collected 31 mins of data from 18 subjects. The data was labelled and engaged/non-engaged and we run a logistic regression in order to find the weights of the features in the model. Our findings show that distance, velocity, body orientation, horizontal head rotation and mouth movement have significant effects on the model. Sound, vertical head orientation and field of view parameters didn't behave as we initially expected. There were also limitations of our work. In order to improve the reliability of the annotations, multiple annotators can be employed by looking at the correlations among the annotators. Instead of using binary engagement labels, annotations can also be done over a range. It will also be interesting to apply and compare other machine learning models. Finally, further data collection and analysis can be done to capture various combinations of user behaviors. Although our results provide useful insights in terms of the importance of feature weights, a validation experiment should be run in order to see whether the new model is more socially adept. This paper is a first attempt to collect and analyze real-life data for engagement detection in a truly open space.

The model is published in the Computer Animation and Virtual Worlds journal and was presented in CASA 2017 [7]. More details about the model and the experiment results can be found in the paper.

2 Description of the Demo

Fig. 2 shows the overall architecture of our system. The Virtual Human Controller receives the information about where to look at from the Engagement Detection component and controls the gaze behavior of the character. The dialogue of the character is based on AIML Pandorabots[1]. For speech recognition, we use Google Speech Recognition.

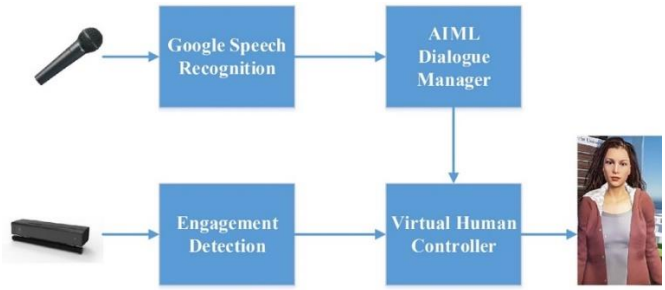

Fig. 2. Overall Architecture

We developed the virtual character in Unity 3D game engine. The 3D model is created in Daz3D[2]. The character has the capability of speaking, gazing, displaying facial expressions, conversational gestures and idle animations. Lip-synch and the low-level gaze movement are based on third-party assets from the Unity Asset Store[3][4]. Gestures are recorded with a Vicon Motion Capture system and applied to our character. Facial expression and visemes are exported as blend shapes from Daz3D. The synchronization between speech, gaze, facial expressions and gestures is realized using the Behavior Mark-up Language [8]. For this, we developed a BML Realizer for Unity[5].

A video of the demo can be found at https://youtu.be/M57kkeoz7zQ. For more information we refer to https://www.staff.science.uu.nl/~yumak001/UUVHC/index.html. Our demo uses two screens. While one screen shows the virtual character, the other one shows the Kinect stream with individual features and overall engagement score in real-time. That enables the visitors to see how the engagement score and the gaze behavior of the character changes based on the underlying model. Fig. 3 shows an example screenshot of the Kinect stream with feature and engagement scores for two people.

[1] http://www.pandorabots.com
[2] http://www.daz3d.com
[3] http://lipsync.rogodigital.com
[4] http://tore-knabe/unity-asset-realistic-eye-movements
[5] https://github.com/christyowidiasmoro/BMLNet

Fig. 3. Features and engagement scores shown on the Kinect stream

References

1. Ruhland, K., Peters, C. E., Andrist, S., Badler, J. B., Badler, N. I., Gleicher, M., Mutlu, B., McDonnell, R.: A review of eye gaze in virtual agents, social robotics and HCI: Behavior generation, user interaction and perception. Computer Graphics Forum, 34(6):299–326, (2015).
2. Sidner, C. L., Lee, C., Kidd, C.D., Lesh, N., Rich, C.: Explorations in engagement for humans and robots. Artificial Intelligence, 166(1- 2):140–164, (2005).
3. Michalowski, M.P., Sabanovic, S., Simmons, R.: A spatial model of engagement for a social robot. In: 9th IEEE International Workshop on Advanced Motion Control, pp. 762–767. IEEE, (2006).
4. Bohus D., Horvitz, E.: Learning to predict engagement with a spoken dialog system in open-world settings. In: Proceedings of the SIGDIAL 2009 Conference, The 10th Annual Meeting of the Special Interest Group on Discourse and Dialogue, pp. 244–252, Stroudsburg, PA, USA, (2009).
5. Foster, M.E., Gaschler, A., Giuliani, M.: How can I help you? Comparing engagement classification strategies for a robot bartender. In: Proceedings of the 15th International Con- ference on Multimodal Interaction (ICMI 2013), Sydney, Australia, (2013).
6. Yumak, Z., Magnenat-Thalmann, N: Multimodal and multi-party social interactions. In: Magnenat-Thalmann, N., Yuan, J., Thalmann, D., You, B. (eds), Context Aware Human-Robot and Human-Agent Interaction, pp. 275–298, Springer International Publishing, (2016)
7. Yumak, Z., van den Brink, B., Egges, A.: Autonomous Social Gaze Model for an Interactive Virtual Character in Real-Life Settings, Computer Animation and Virtual Worlds, (2017).
8. Kopp, S., Krenn, B., Marsella, S., Marshall, A. N., Pelachaud, C., Pirker, H., Thorisson, K.R., Vilhjalmsson, H.: Towards a common framework for multi-modal generation: The behavior markup language. In: Proceedings of the 6th International Conference on Intelligent Virtual Agents, IVA'06, pp. 205–217, Springer-Verlag, Berlin, Heidelberg, (2006).