In Proceedings of Intelligent Virtual Agents (IVA), Reykjavík, Iceland, 2011

Modeling Gaze Behavior for Virtual Demonstrators

Yazhou Huang, Justin L. Matthews, Teenie Matlock, and Marcelo Kallmann

University of California, Merced

Abstract. Achieving autonomous virtual humans with coherent and natural motions is key for being effective in many educational, training and therapeutic applications. Among several aspects to be considered, the gaze behavior is an important non-verbal communication channel that plays a vital role in the effectiveness of the obtained animations. This paper focuses on analyzing gaze behavior in demonstrative tasks involving arbitrary locations for target objects and listeners. Our analysis is based on full-body motions captured from human participants performing real demonstrative tasks in varied situations. We address temporal information and coordination with targets and observers at varied positions.

Keywords: gaze model, motion synthesis, virtual humans, virtual reality.

1 Introduction and Related Work

Human-human interactions are ubiquitous and in some cases necessary for survival. Engaging in joint activities, such as working on a text together, discussing dinner plans, or showing a friend where to park a car with pointing gestures seem trivial and effortless. However such interactions are orchestrated with a high level of complexity. They may consist of multiple levels of coordination, from conversational communication to gesture, and to the combination of speech and gesture [3, 10]. A good understanding and modeling of these multiple levels of coordinated language and action can help guide the design and development of effective intelligent virtual agents. An important part of this is the study of gaze behavior.

The immediate goal of our work is to generate humanlike full-body motions that are effective for demonstration of physical actions to human users by means of a virtual character. In our approach, the virtual trainer has also to position itself in a suitable location for the demonstration task at hand. This is in particular important to guarantee that the actions and target objects are visible to the observer. Gestures and actions need to be executed by the virtual demonstrator with clarity and precision in order to appropriately reference the target objects without ambiguity. Human users are highly sensitive to momentary multi-modal behaviors generated by virtual agents [20]. In addition, the speed of the motion in the articulation of such behaviors is important in the use and understanding of manual movement [8]. This paper presents our first results analyzing gaze behavior and body positioning for a virtual character identifying and delivering information about objects to an observer in varied relative locations.

In order to investigate these issues we have conducted several motion capture sessions of human-human demonstrative tasks. The collected data is full-body and reveals important correlations that can be directly integrated into gaze and body coordination models for virtual humans. Our results are being integrated in our training

framework [2] based on virtual agents that can learn clusters of demonstrative gestures and actions [7] through an immersive motion capture interface.

There is a large body of research on modeling gaze in humans and animals. Some of this neurological research focuses on the nature of eye movements, including saccades (ballistic eye movements that jump from location to location in a visual scene in a matter of milliseconds) [15, 17]. Some studies [12] closely examine vestibulo-ocular (VOR) reflex in saccadic and slow phase components of gaze shifts. Additional studies [4, 6] involve fine-grained analysis small and large gaze shifts where classic feedback loops are used to model the coupling and dynamics of eye and head-orienting movements.

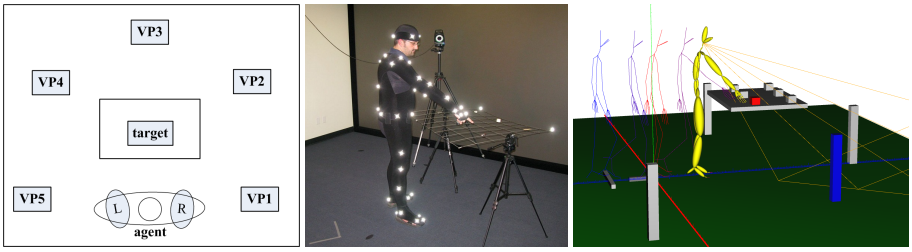
Gaze has been used in computer graphics from gaze-contingent real-time level of detail (LOD) rendering [13] to the modeling of movement for eyes balls, eye lids and related facial expressions [5]. Gaze direction in particular is known to help with basic two-way communication because it can help a speaker direct attention and disambiguate for a listener [9]. Gaze direction has also been shown to help human listeners better memorize and recall information in interactions with humanoid interlocutors, including robot storytellers [14] or a narrative virtual agent in a CAVE system [1]. [11, 16] introduce emotion models with body posture control to make synthesized gaze emotionally expressive. These systems typically use pre-recorded voice coupled with simulated gaze to interact with the listener. The controlled agent will remain in the same spot facing the audience, and without the need for locomotion.

In this paper we analyze higher-level gaze behavior together with important gaze-related events such as body positioning, synchronization with pointing gestures in respect to multiple objects in the workspace, and with the purpose of delivering information to a human observer at different locations.

2 Data Collection Setup

A total of 4 male participants (weight $150 \sim 230$ lb, height $5'9 \sim 6'1$) were recruited to perform a variety of basic pointing tasks with full-body motion capture without eye tracking. The capture environment was an 8 foot x 12 foot rectangle area. It included six small target objects (office supplies) that were placed on a horizontal coarse mesh grid (simulating a table). Each participant's action was observed by a viewer (human observer) standing at viewer's perspective (VP) locations VP1 though VP5, see Figure 1 (a) and (b). In order to avoid possible effects of target size on gaze behavior, small targets were specifically selected.

For each trial of the motion capture, the participant (1) stands about 4 feet away from the mesh grid, (2) walks towards the grid, (3) points to one of the target objects, (4) verbally engages with the viewer by either naming the target ("This is a roll of tape"), physically describes it ("small, smooth, and black"), or describes the function ("It's used for holding things in place"). During each trial, the participant is expected to direct the attention of the viewer as needed while pointing and talking, by naturally gazing back and forth at the viewer and target. The participant then steps back to the starting position and prepares for the next trial. Each capture session includes 30 trials. The viewer maintains the observing position until all 6 targets had been addressed, then moves to the next standing location. This sequence is repeated until all targets



(a) (b) (c)

Fig. 1. (a) and (b): motion capture setup; (c): a snapshot of our annotation application showing the phase-plane of gaze yaw-axis along with reconstructed environment.

are named or described to the viewer at each of the 5 VPs. The sequence of target selections was random. The full-body motion data (without eye tracking) was captured at 120 fps then retargeted and down-sampled to 60 fps. Data was annotated manually using our annotation tool (Figure 1 (c)). Each captured sequence contains many streams of information, in the present work we have annotated the motions with the information relevant for analyzing the observed gaze behavior.

3 Analysis and Discussion

Our first observation is that each trial was typically constituted of a series of largely consistent gaze or gaze-related events, as listed below:

1. the participant gazes at the floor when walking towards the target object;
2. the participant gazes at the target to be addressed with the demonstrative action;
3. stroke point of the action, in the case of pointing, is detected by the zero-crossing frame of the velocity of the participant's end-effector (the hand);
4. the participant gazes at the viewer during the action and while describing the target;
5. the participant again gazes at the target during action, if applicable;
6. the participant again gazes at the viewer during action, if applicable;
7. the participant gazes at any additional (irrelevant) locations, if applicable;
8. the participant gazes at the floor when stepping back to initial location.

Annotations were then performed to precisely mark the time stamps (start/end) of each event listed above. In the next sections we interpret the annotated events in respect to (a) temporal parameters related to gaze behavior and (b) gaze-related body positioning patterns for demonstrative tasks.

3.1 Temporal Parameters for Gaze Behavior Modeling

The first analysis focuses on the temporal delay Δt between the action stroke point and the starting of the gaze-at-viewer event. Annotation results show that when the viewer is positioned within participant's field-of-view (FoV) (i.e. VP2, VP3, VP4 in Fig 1(a)),

the gaze-at-viewer event immediately follows the pointing action stroke point, resulting in $\Delta t > 0$. By contrast, when the viewer is outside of FoV (i.e. VP1 and VP5 in Fig 1(a)), due to the large gaze-shift required to visually engage with the viewer, gaze-at-viewer starts ahead of the action stroke point, and in this case $\Delta t < 0$. This temporal delay extracted from the trials (measured in seconds) is plotted in Figure 2.

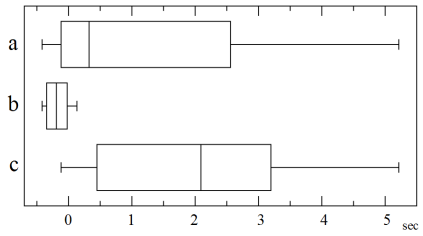


Fig. 2. Temporal delay dictates the starting time of gaze-at-viewer before or after action stroke point: (a) delay plot across all trials. (b) delay plot for out-of-FoV viewer positions VP1 and VP5. (c) delay plot for inside-FoV viewer positions VP2, VP3 and VP4.

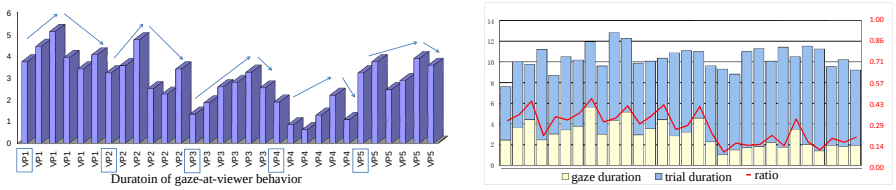
The second analysis reveals correlations between gaze-at-viewer durations and viewer positions. During the capture sessions the viewer moves to a new position after the participant addresses all 6 target objects on the table. An interesting pattern over the gaze-at-viewer durations can be observed across all participants: the viewer switching to a new position results in an increase in the gaze duration, which typically lasts for 2 to 4 trials. This increase is shortly followed by gradual declines in gaze duration, see Figure 3(a). Studies from psychological research on animals resonates to this result [18], specifically when the declination of responsive behavior in humans (extinction progress) begins, a brief surge often occurs in the responding, followed by a gradual decline in response rate until it approaches zero.

The third analysis focuses on the gradual decline of gaze-at-viewer durations. The duration each participant takes to verbally name and describe each object varies across trials. To discount such variation, the ratio (percentage) of gaze-at-viewer behavior takes up within each trial is observed, see Figure 3(b). Dark bars and clear bars correspond to durations of the trial and of the gaze behavior, respectively. Red line drawing reflects the aforementioned ratio decline.

Lastly, to generate natural head movements for gaze behavior, a velocity profile similar to [19] is used to dictate humanlike head rotations based on angular acceleration/deceleration patterns from captured data, see Figure 4 (unfiltered raw data plot).

3.2 Gaze-Related Body Positioning Patterns

Body positioning is of great importance and is rarely addressed. The positioning of a virtual trainer is critical for the viewer to have clear understanding of the action being demonstrated. It is of the same importance to the virtual trainer so that natural gaze behaviors can be carried out visually engaging with the viewer.



(a) (b)

Fig. 3. (a) Correlations between gaze-at-viewer durations and viewer positions: when viewer switches to a new position (boxed text), gaze duration increases for $2 \sim 4$ subsequent trials, then declines. (b) Gradual decline of gaze-at-viewer durations over time. Lighter vertical bars: gaze duration; Darker bars: trial duration; line graph: ratio of gaze duration over trial duration.

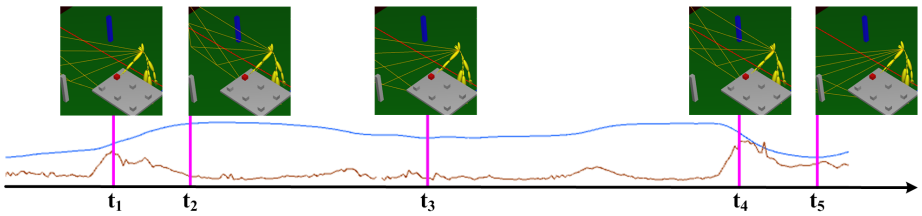


Fig. 4. The velocity profile observed in head motions from captured gaze behavior (unfiltered). The lower trajectory reflects angular accelerations/decelerations of the head rotations, which are used to simulate head movement for gaze. The upper bell-shaped line measures the angle in head rotations from the rest posture (head looking forward). t_1: start of gaze-at-viewer; t_2: gazing at viewer; t_3: gaze-at-target during describing the target; t_4: end of gaze-at-viewer; t_5: start of another gaze-at-target.

We have extracted from the captured data (from one participant) the parameters defined in Figure 5, and their values are summarized in Table 1. The dashed reference line is perpendicular to the table edge, from which the agent will approach the table. In respect to the target object on the table, α and β measures the relative standing locations for the viewer and the demonstrative agent respectively. β dictates where the agent positions itself giving the viewer a clear view at the target; θ dictates how the agent orients its body to conduct demonstrative actions towards the viewer. ϕ is the recorded maximum head rotation (gaze shift) during the gaze-at-viewer behavior. For any new environment, only α will be treated as an input value, while β, θ and ϕ need to be learned from captured data to solve the gaze-related body positioning problem.

4 Conclusion

In this paper we have discussed studies used for analyzing and modeling gaze behavior for virtual trainers performing object demonstrations. Several aspects of the collected

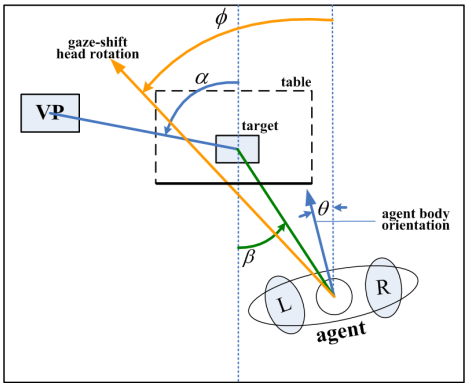


Fig. 5. Description of key gaze-related body positioning parameters: α and β are relative standing locations for the viewer and the agent respectively. θ dictates body orientation of the agent, ϕ represents the maximum head rotation in gaze-at-viewer behavior.

setup	body positioning parameters							
VP	$\bar{\alpha}$	α_α	$\bar{\beta}$	α_β	$\bar{\theta}$	α_θ	$\bar{\phi}$	α_ϕ
VP1	129.0	4.8	-20.6	10.1	-25.1	7.4	-82.7	4.6
VP2	76.6	8.2	-15.9	10.4	-4.8	5.8	-32.2	4.9
VP3	-2.8	14.3	-7.3	11.2	11.6	9.1	0.7	5.1
VP4	93.8	6.3	13.7	12.7	44.5	8.9	58.4	2.7
VP5	149.2	5.5	24.0	11.8	76.0	6.8	125.8	4.0

Table 1. Body positioning parameters observed from one participant performing the action towards different targets and viewer positions. For each parameter, the first column shows the average value (each computed from 6 trials with the viewer maintaining its position), and the second column is the coresponding average absolute deviation of first column, in degrees.

full-body motion data were analyzed in respect to gaze behaviors and gaze-related body-positioning. Our first results presented in this paper lead to several informative correlations for implementing animation models for controlling virtual humans in interactive training systems. In future work we will present a comprehensive analysis of the entire motion information collected, and we will present complete behavioral models for realistically animating full-body virtual trainers in demonstration scenarios.

Acknowledgements This work was partially supported by NSF Awards IIS-0915665 and CNS-0723281. The authors would like to thank David Sparks for his assistance in motion annotations, and all the participants involved in the data collection process.

References

1. Bee, N., Wagner, J., André, E., Vogt, T., Charles, F., Pizzi, D., Cavazza, M.: Discovering eye gaze behavior during human-agent conversation in an interactive storytelling applica-

tion. In: Int'l Conference on Multimodal Interfaces and Workshop on Machine Learning for Multimodal Interaction. pp. 9:1–9:8. ICMI-MLMI '10, ACM, New York, NY, USA (2010)
2. Camporesi, C., Huang, Y., Kallmann, M.: Interactive motion modeling and parameterization by direct demonstration. In: Proceedings of the 10th International Conference on Intelligent Virtual Agents (IVA) (2010)
3. Clark, H.H., Krych, M.A.: Speaking while monitoring addressees for understanding. Memory and Language 50, 62–81 (2004)
4. Cullen, K.E., Huterer, M., Braidwood, D.A., Sylvestre, P.A.: Time course of vestibuloocular reflex suppression during gaze shifts. Journal of Neurophysiology 92(6), 3408–3422 (2004)
5. Deng, Z., Lewis, J., Neumann, U.: Automated eye motion using texture synthesis. Computer Graphics and Applications, IEEE 25(2), 24 – 30 (2005)
6. Galiana, H.L., Guitton, D.: Central organization and modeling of eye-head coordination during orienting gaze shifts. Annals of the New York Acd. of Sci. 656(1), 452–471 (1992)
7. Huang, Y., Kallmann, M.: Motion parameterization with inverse blending. In: Proceedings of the Third International Conference on Motion In Games. Springer, Berlin (2010)
8. Huette, S., Huang, Y., Kallmann, M., Matlock, T., Matthews, J.L.: Gesture variants and cognitive constraints for interactive virtual reality training systems. In: Proceeding of 16th International Conference on Intelligent User Interfaces (IUI). pp. 351–354 (2011)
9. Kendon, A.: Some functions of gaze direction in two-person conversation. Conducting Interaction: Patterns of Behavior in Focused Encounters (1990)
10. Kendon, A.: ADAM KENDON, Gesture: Visible action as utterance. Cambridge (2004)
11. Lance, B., Marsella, S.: Emotionally Expressive Head and Body Movements During Gaze Shifts. In: 7th Int'l Conference on Intelligent Virtual Agemts (IVA). pp. 72–85 (2007)
12. Lefevre, P., Bottemanne, I., Roucoux, A.: Experimental study and modeling of vestibulo-ocular reflex modulation during large shifts of gaze in humans. Experimental Brain Research 91, 496–508 (1992)
13. Murphy, H.A., Duchowski, A.T., Tyrrell, R.A.: Hybrid image/model-based gaze-contingent rendering. ACM Trans. Appl. Percept. 5, 22:1–22:21 (February 2009)
14. Mutlu, B., Hodgins, J.K., Forlizzi, J.: A storytelling robot: Modeling and evaluation of human-like gaze behavior. In: Proceedings of HUMANOIDS'06, 2006 IEEE-RAS International Conference on Humanoid Robots. IEEE (December 2006)
15. Pelisson, D., Prablanc, C., Urquizar, C.: Vestibuloocular reflex inhibition and gaze saccade control characteristics during eye-head orientation in humans. Journal of Neurophysiology 59, 997–1013 (1988)
16. Thiebaux, M., Lance, B., Marsella, S.: Real-time expressive gaze animation for virtual humans. In: Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems (AAMAS). pp. 321–328 (2009)
17. Van Horn, M.R., Sylvestre, P.A., Cullen, K.E.: The brain stem saccadic burst generator encodes gaze in three-dimensional space. J. of Neurophysiology 99(5), 2602–2616 (2008)
18. Weiten, W.: Wayne Weiten, Psychology: Themes and Variations. Cengage Learning Publishing, 8th edition (2008)
19. Yamane, K., Kuffner, J.J., Hodgins, J.K.: Synthesizing animations of human manipulation tasks. In: ACM SIGGRAPH 2004. pp. 532–539. ACM (2004)
20. Zhang, H., Fricker, D., Smith, T.G., Yu, C.: Real-time adaptive behaviors in multimodal human-avatar interactions. In: Int'l Conf. on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction. pp. 4:1–4:8. ICMI-MLMI '10, ACM (2010)