Journal of Universal Computer Science, vol. 13, no. 2 (2007), 149-160
submitted: 31/7/06, accepted: 15/1/07, appeared: 28/2/07 © J.UCS

Building Immersive Conversation Environment Using Locomotive Interactive Character

Rai Chan
(University of Tsukuba, Japan
zhanglei@graphic.esys.tsukuba.ac.jp)

Junichi Hoshino
(University of Tsukuba, Japan
jhoshino@esys.tsukuba.ac.jp)

Abstract: Generating composite human motion such as locomotion and gestures is important for interactive applications, such as interactive storytelling and computer games. In interactive story environments, CG characters do not merely stand in one position. Rather, they should be able to compose gestures and locomotion based on the discourse of the story and the locations of objects in the scene. Thus, in the present paper, we propose a conversational locomotion model for CG character. We constructed a conversational locomotion network for a virtual environment. A multi-path searching algorithm calculates the optimal walking path, which uses node activation from the story locations and conversation units. The CG character also locally adjusts its position so that it does not block the referenced object from the user's sight. We have applied the proposed technique to an interactive 3D movie system and have demonstrated composite motion of the locomotion and conversation of a CG character, which improves the immersion of the viewer in the story environment.

Keywords: conversational locomotion, conversational locomotion network, panorama
Categories: D.1.0, D.1.1, D.2.1

1 Introduction

In daily life, humans perform several composite actions simultaneously. Walking and talking concurrently is one typical composite human action. Composing locomotion and gestures is also important for applications such as interactive movies and games. In the interactive story environment, CG characters do not simply stand in one position, they should be able to compose gestures and locomotion based on the story location and surrounding objects.

The proper location and timing of the CG character is influenced by various contexts, such as the connection of scene locations and the current environment. The apparent size of objects and the degree of detail of the explanation affect how much closer the CG character should move to a referenced object. Connection of the scene locations also affects the current position. It is time consuming to have the CG character approach the object each time that a CG character refers to a particular object during a conversation when the object is far from the CG character's current position. However, when the referenced object is close to the next scene location, it is reasonable that the CG character should move closer to the object.

Several research papers have discussed the building of embodied conversational agents with various degrees of conversational ability. [Ball et al. 97] are building an embodied conversational interface that will eventually integrate spoken language input, a conversational dialogue manager, reactive 3D animation, and recorded speech output. [Loyall and Bates 97] developed engaging characters that allow the viewer to suspend disbelief long enough to interact in interesting ways with the character, or to be engaged by the interactions of the character with another computer character. [Noser et al. 95] presented a navigation model for animation characters using synthetic vision. [Bandi and Thalmann 98] discretized a synthetic environment into a 3D uniform grid in order to search paths for autonomous characters. [Rose et al. 98] introduced the framework of “verbs and adverbs” to interpolate example motions with a combination of radial basis functions.

In the present paper, we propose a mechanism for fluid conversational locomotion for a CG character. This mechanism is realized by calculating the optimal locomotion path, which is influenced by the conversation and the story location, and the CG characters subsequently generate composite walking and conversation actions. The CG character also locally adjusts its position so as not to block the referenced object from the user’s sight.

This technology would contribute to interactive edutainment systems. For example, in order to gain knowledge of disaster prevention in large-scale disasters such as the Hanshin-Awaji Earthquake and the Niigata-ken Chuetsu Earthquake, the user must recognize the surrounding situation by moving and taking appropriate action based on the advice of a CG character. Figure 1 shows a typical example of a locomotive conversation. In this scenario, the character first communicates disaster prevention knowledge, stating that vending machines in the disaster zone should not be approached. In the next scene, the user asks about other specific objects that should not be approached, and the character then moves closer to explain more about the background of the disaster.

Figure 1: Example of conversational locomotion

2 Synchronization of Conversation and Locomotion

2.1 Overview of the architecture

Figure 2 shows the conversational locomotion architecture. The system has a locomotion module, conversation modules, and a story manager. A story consists of a set of scene units, and this controls the discourse of the conversation. A scene unit has a precondition, scene location, and links to a collection of possible conversation modules. Proper scene units are selected using the preconditions, such as changes in the environment and the history of the user's verbal expressions. When a story unit is selected, possible conversation units applied in the scenes are activated.

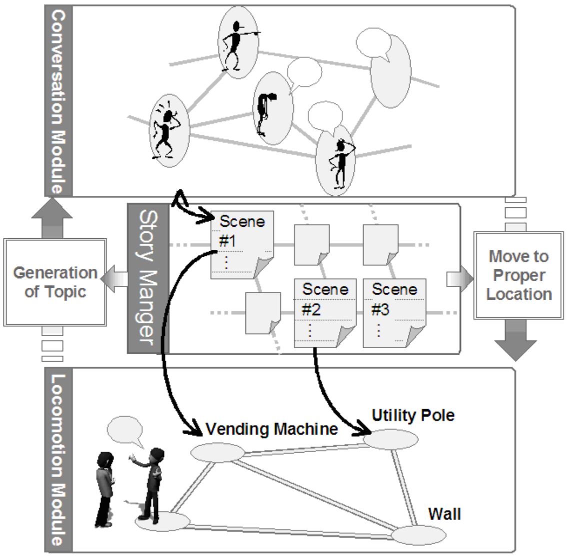
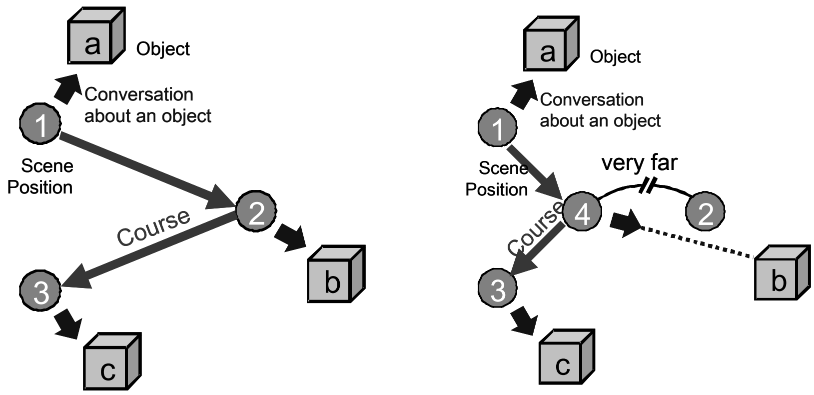


Figure 2: Overview of the conversational locomotion architecture

Conversation modules have preconditions, utterances, corresponding gestures and key locations. The locomotion module dynamically plans locomotion paths and generates walking motion patterns based on key locations.

2.2 Key location control

To compose locomotion and conversation, we need to decide the CG character's location and the timing of walking during the conversation. The locations of the CG character are influenced by where the scenes are taking place and the content of the conversations. Figure 3 shows a typical example of locomotion planning during a conversation. In Figure 3(a), there are scene positions node 1, node 2, and node 3. Assuming that the present location is node 1, after the CG character talks about object "b" at node 2, it should move to node 3, and the movement path corresponding to the scene position is indicated by a red arrow. However, In Figure 3(b), when node 2 is very far and object "b" is sufficiently visible from node 4, it is reasonable for the CG character to stop at node 4 to talk about object "b".



a) When location (2) is near b) When location (2) is far

Figure 3: Concept of conversational locomotion using a simple example

Locomotion and conversation are composed by considering the following three types of location constraints:

Scene location: The scene location corresponds to where the actions and conversations are taking place. To begin a conversational scene, the CG character should be at the proper location.

Interpersonal location: The CG character changes relative locations with respect to other CG characters during conversation. For example, when the CG character begins to talk, the CG character approaches the other participants. When the CG character tries to explain something, the visibility of the referenced object is also considered when deciding interpersonal location.

Reference location: This is the relative location of the CG character and the referenced object.

These location constraints are used as key locations k_n in the conversational locomotion planning. The key location consists of a position in the floor coordinate system, and a standing duration t_k at a given key location. In most scenes, the proper standing position of the CG character has a degree of freedom. A key location has a several candidate positions with different activation values.

The standing duration of the key location can be dynamically changed by the key location control rules in the conversation units. For example, the initial standing duration can be used to decide how long the CG character can talk with the user at a particular position. When a conversation with the user ends, the conversation units set the standing duration to zero, which then causes the CG character to move onto the next scene location.

2.3 Conversational locomotion network

Activating the locomotion network using story locations and conversation units controls conversational locomotion. An optimal locomotion path is selected by calculating the optimal locomotion path with the maximum activation.

Locomotion node k represents a point in the floor coordinate system (u_k, v_k). To identify each 3D position by nodes, we assign a tag with a unique name to each position, such as "Vending Machine" or "Utility Pole". The CG character can walk away from locomotion nodes for local position adjustment.

In the locomotion network, we define *K* as the position set and *E* as the node set that expresses transitions between nodes. The directed graph is then given as $G_k = (K, E)$. The distance of the arc $a = (u_k, v_k)$ that connects adjacent locations u_k and v_k is expressed as $distance_k(a)$, and the locomotion network is $N_k = (G_k, distance_k)$.

The initial locomotion network is constructed by sampling the possible standing locations. The candidate node positions are story locations and objects that have been referenced previously in conversation units. To increase the number of possible locations, we randomly sample the possible walking space.

Associating a key location with a proper clause in an utterance controls the timing of locomotion. For example, the reference location can be associated with clauses including the referenced object. There are several methods for associating the key location with a clause. When the numbers of conversational modules are limited manual association may be relatively easy. Even if the key location specification is predetermined, the CG character motion is changed dynamically depending on the story locations and the order of the conversations.

The timing of utterances is also synchronized with CG character locomotion. For example, the preconditions of conversation units are used for pauses and thereby presentation until the CG character moves to the proper positions.

3 Conversational Locomotion Planning

Key locations are activated using scene locations and activation rules in the conversation modules. The locomotion path is selected dynamically using a multi-path searching algorithm that calculates the maximum activated path. When the conversation units change the status of activation, the locomotion path is recalculated.

3.1 Multiple path searching

Multiple key locations are set with different activation values. By selecting N-best key locations, the possible locomotion segments between key locations are selected. The total activation along the locomotion segments is calculated. In addition, to calculate the shortest path for the locomotion, the locomotion segments between candidate key locations are obtained by employing the method of Dijkstra described previously [Dijkstra 59]. Although we used the Dijkstra algorithm in the present paper, in the future, in order to reduce the time required for calculation, we will consider optimizing the algorithm by emulating the Dijkstra algorithm [Tsitsiklis 95]. As shown in Figure 4, we calculated candidate locomotion segments, such as P_{01} - P_{11}

and P_{01} - P_{12}, to obtain the total activation value, and the path with the maximum activation value is assumed to be CG character's movement path.

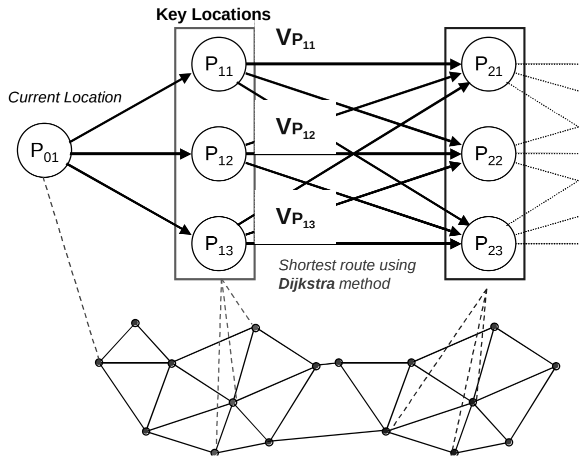
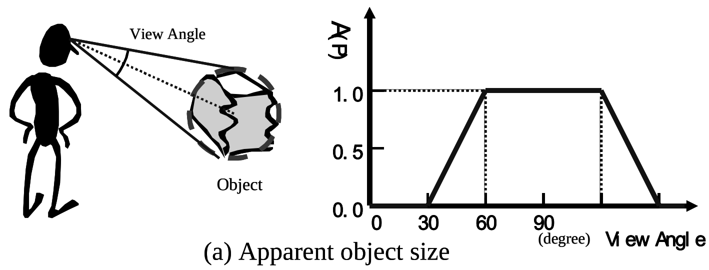


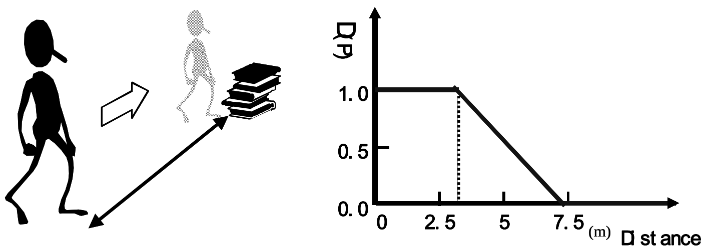
Figure 4: Key location and multiple path searching

3.2 Activation functions

In addition to the scene locations, we use the apparent object size and walking size to locally control locomotion. Figures 5(a) and 5(b) show the activation function used in this system.



(a) Apparent object size



(b) Distance from the object

Figure 5: Scene constraints and activation value

(1) Apparent object size: $A(P_{s,n})$

When the CG character talks about a referenced object, it should stand a certain distance from the object when providing the introductory explanation and should move closer to the object when providing a more detailed explanation. We determined the activation function as shown in Figure 5(a). A sphere approximates the reference object, and the view angle from the CG character's eye position is calculated. Note that the approximated object size corresponds to the object area referred to in the conversation. When the CG character refers to part of a large object, the approximated object size is small. Orientation constraints are also integrated by generating the activation distribution in a specific direction.

In the present paper, the angle at which the object is located in a stable gaze field $(60^o \le \theta \le 120^o)$ is a threshold, and the activation value is decided as follows:

$$A(P_{s,n}) = \begin{cases} (\theta - 30)/30 & (30° \le \theta < 60°) \\ 1.0 & (60° \le \theta < 120°) \\ -(\theta - 120)/30 + 1.0 & (120° \le \theta < 150°) \\ 0.0 & (otherwize) \end{cases} \quad (1)$$

where $A(P_{s,n})$ is the activation value at position $P_{s,n}$, and θ is the viewing angle at position $P_{s,n}$.

(2) Walking distance: $D(P_{s,n}, P_{s+1,m})$

When the walking distance from the current location of the CG character is longer, the CG character tries to avoid this longer path. Therefore, while the distance is short, the activation value of the walking distance is high. Considering the normal movement range that humans use when explaining objects in daily life, in the present paper, we set the distance that moving easily is 0 ~ 3 m, and the distance that possible to move is 3 ~ 7.5 m. Because the user will not want to move great distances, the linear change was thought happened in the possible moving range. We then determined the activation function as depicted in Figure 5(b). The activation value is decided as follows:

$$D(P_{s,n}, P_{s+1,m}) = \begin{cases} 1.0 & (0[m] \le d < 3.0[m]) \\ (3.0 - d)/4.5 + 1.0 & (3.0[m] \le d < 7.5[m]) \\ 0.0 & (otherwize) \end{cases} \quad (2)$$

where $D(P_{s,n}, P_{s+1,m})$ is the activation value between position $P_{s,n}$ and position $P_{s+1,m}$, and d is the distance between position $P_{s,n}$ and position $P_{s+1,m}$.

Based on Eqs. 1 and 2, the total activation values of each movement path $P_{0,0}, P_{1,n_1}, ..., P_{s,n_s}$ are calculated as follows:

$$V(P_{0,0}, P_{1,n_1}, ..., P_{s,n_s}) = \sum_{t=1}^{s} \{w(t) \cdot [\alpha A(P_{t,n_t}) + (1-\alpha) D(P_{t-1,n_{t-1}}, P_{t,n_t})]\} \quad (3)$$

where $w(t)$ is a weighting value that is used to control the number of key locations that the CG character should consider, and α ($0 \le \alpha \le 1$) is the sharing rate of apparent object size and walking distance.

Another type of activation function is easily integrated into this framework. For example, access control of the CG character to a specific area can be represented.

Assigning a negative activation value to specific locations will cause the CG character to avoid entering the area.

3.3 Local position adjustment

The CG character's position is locally adjusted so as not to obscure the user's view of the referenced object. Figure 6 shows the concept of local position adjustment. As described in Section 3.2, we approximate the reference area using a sphere, and an isosceles triangle formed by the user's aspect and the diameter of object is the object gaze area. The viewing area of the user is calculated from the 3D location of the user's view camera and the reference object sphere. When the CG character approaches an object, it stops at the intersection of the view area and the edges of the locomotion network.

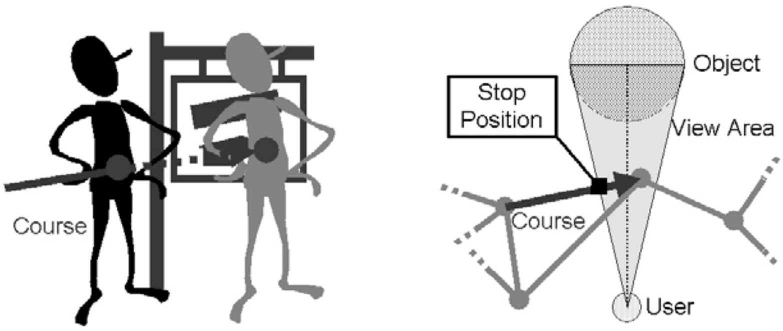


Figure 6: Local position adjustment

4 Panorama-based Immersive Story Environment

4.1 Introduction

The construction of convincing photo-real 3D models is very time consuming. As such, we first attempted to build a panorama-based immersive story environment to prove our theory. CG characters can walk and talk in a photo realistic environment by using a combination of locomotion network and object annotations. A model of the environment is approximated by the linked panoramic images. The CG character and the user can thus move around within the photo-real scene.

4.2 Generation of node coordinates in 3D space

To facilitate walking of the CG character in the environment, a correspondence between the 2D panoramic image and its 3D environment is needed. We constructed a locomotion network to determine the walking area and calculate the appropriate walking path. Figure 7b shows an example of the locomotion network. In order to avoid approaching an object that is designated as being dangerous, the locomotion area must be indicated to the CG character. Therefore, the locomotive area that can be passed through is specified with the paint by user beforehand. Then, the nodes that show the CG character can arrive are automatically generated by the painted area.

Finally, the 3D locations of scene objects are annotated so that the CG character can refer to and point at these objects during conversations.

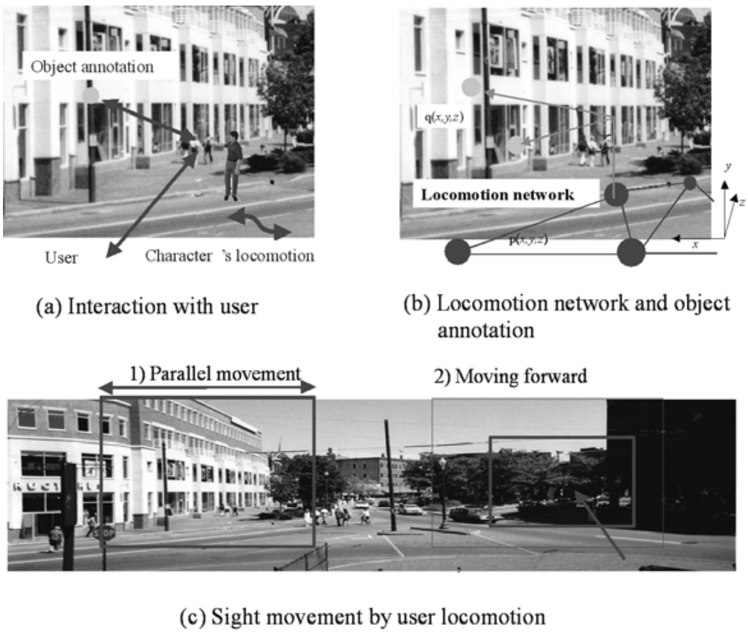


Figure 7: Interaction mechanism for panorama-based conversation environment

When the camera coordinates $\mathbf{P}_c = (x_c, y_c, z_c)$ and the 2D coordinates of the node in the panoramic image are assumed to be $\mathbf{P}_n^{2D} = (x_n^{2D}, y_n^{2D})$, the 3D space in which a virtual CG character exists $\mathbf{P}_p^{3D} = (x_p^{3D}, y_p^{3D}, z_p^{3D})$ is converted into the projection coordinates by

$$x_p^{3D} = x_c^{3D} \quad (4)$$

$$y_p^{3D} = \frac{2C_y}{H} y_n^{2D} \quad (5)$$

$$z_p^{3D} = f \quad (6)$$

where f is the distance from the camera to the plane of projection, C_y is the proportion of the height of the panoramic image to the plane of projection, and H is the height of the panoramic image.

When a CG character is displayed in a panoramic image, it is necessary to convert the projection coordinates $\mathbf{P}_p^{3D}$ into the coordinates $\mathbf{P}_n^{3D}$ of the 3D space in order to present the image that moves in the direction as depth changes.

The straight-line equation extracted from the camera coordinates $\mathbf{P}_c$ to the projection coordinates $\mathbf{P}_p^{3D}$ is as follows:

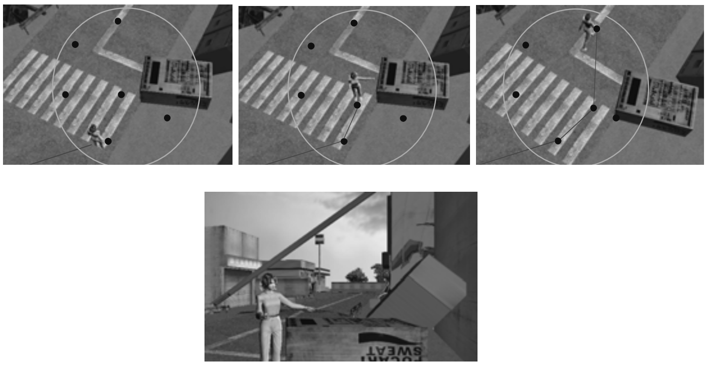
$$y = \frac{y_p^{3D}}{f}(z - z_c) + y_c \,. \quad (7)$$

By employing Eq. (7), when either the z value or the y value is determined, the other value is also determined. Then, by the following equation, the z value is set to be proportional to the height in the panoramic coordinates:

$$z_n^{3D} = \left(z_{\max}^{3D} - z_{\min}^{3D}\right)\frac{\left(y_n^{2D} + \frac{H}{2}\right)}{y_{\max}^{2D}} + z_{\min}^{3D} \cdot \qquad (8)$$

5 Results

Furthermore, we have applied the proposed composite motion generation technique to interactive 3D movie applications. As mentioned earlier, in the interactive story environment CG characters do not simply stand in one position, because they constantly interact with their environment. If a CG character changes its responses depending on the user's input, the CG character should then decide when and where to move.



(a) Conversational locomotion with path planning

(b) Conversational locomotion without path planning

Figure 8: Screenshots from a conversational locomotion sequence

To develop this application, we chose the C++ language for system development and used OpenGL to provide 3D views of the environment. The entire system was implemented on an Intel 1-GHz Pentium III computer running the Windows XP operating system.

5.1 Composition of locomotion and conversation

In this experiment, a model of a virtual environment was constructed using 3D studio Max 5.0, and a CG character was created using Poser 5.0. Figure 8 shows the results of the conversational locomotion. To synchronize locomotion and conversation, we need to determine the location and the timing of locomotion during conversation. Figure 8(a) shows a screenshot with locomotion planning, while Figure 8(b) shows a screenshot without locomotion planning. The CG character tries to select a closer position to explain relevant details.

5.2 Locomotion conversation based on a panoramic image

In this experiment, the CG character was created using Haptek Inc.'s People Putty [Haptek], and a conversation scene of a panoramic image that includes a church (Figure 9) was constructed to confirm the effectiveness of the technique. In this panoramic image, the technique mentioned previously was applied, and a CG character moved in the panoramic space while simultaneously talking with the user. Thus, a movie in which a CG character acted as a guide was generated very effectively (Figure 10).

Figure 9: Panoramic image of a church

Figure 10: Screenshots from virtual trip sequences

6 Conclusion

In the present paper, we have proposed a conversational locomotion model for a CG character. By calculating the optimal locomotion path influenced by both conversation and story location, the CG character generates composite walking and conversation actions. The CG character also locally adjusts the position considering the visibility of relevant objects with respect to the user. As a result, we applied this technique to educational applications and produced a highly accurate interactive animation sequence using conversational locomotion. We believe that the application of the proposed technique will significantly improve current interactive educational applications.

References

[Ball et al. 97] Ball G., Ling D., Kurlander D., Miller J., Pugh D., Skelly T., Stankosky A., Thiel D., Van Dantzich M., and Wax T.: “Lifelike computer characters: the persona project at Microsoft Research”; Proc. Software Agents Publishing, Cambridge (1997), 191-222.

[Bandi and Thalmann 98] Bandi S. and Thalmann D.: “Space Discretization for Efficient Human Navigation”; Proc. Eurographics’98, Computer Graphics Forum Publishing (1998), 195-206.

[Dijkstra 59] Dijkstra E.W.: “A note on two problems in connection with graphs”; Proc. Numerische Mathematik Publishing (1959), 269-271.

[Haptek] http://www.haptek.com

[Loyall and Bates 97] Loyall A.B. and Bates J.: “Personality-rich believable agents that use language”; Proc. The First International Conference on Autonomous Agents Publishing (1997), 106-113.

[Noser et al. 95] Noser H., Renault O., Thalmann D. and Thalmann N.M.: “Navigation for Digital Actors Based on Synthetic Vision, Memory and Learning”; Proc. Computers and Graphics Publishing (1995), 7-19.

[Rose et al. 98] Rose C., Bodenheimer B. and Cohen M.: “Verbs and Adverbs: Multidimensional motion interpolation using radial basis functions”; Proc. IEEE Computer Graphics and Applications Publishing (1998), 32-40.

[Tsitsiklis 95] Tsitsiklis J.N.: “Efficient Algorithms for Globally Optimal Trajectories”; Proc. IEEE Transactions on Automatic Control Publishing (1995), 1528-1538.