World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:3, No:11, 2009

# Humanoid Personalized Avatar Through Multiple Natural Language Processing

Jin Hou[*ab], Xia Wang[a], Fang Xu[a], Viet Dung Nguyen[a], Ling Wu[a]
[a]School of Information Science and Technology, Southwest Jiaotong University,
Chengdu, Sichuan 610031, P.R. China
jhou@home.swjtu.edu.cn
[b]State Key Laboratory for Novel Software Technology, Nanjing University,
Nanjing, Jiangsu 210093, P.R. China

Open Science Index, Computer and Information Engineering Vol:3, No:11, 2009 publications.waset.org/13837/pdf

***Abstract*—**There has been a growing interest in implementing humanoid avatars in networked virtual environment. However, most existing avatar communication systems do not take avatars' social backgrounds into consideration. This paper proposes a novel humanoid avatar animation system to represent personalities and facial emotions of avatars based on culture, profession, mood, age, taste, and so forth. We extract semantic keywords from the input text through natural language processing, and then the animations of personalized avatars are retrieved and displayed according to the order of the keywords. Our primary work is focused on giving avatars run-time instruction from multiple natural languages. Experiments with Chinese, Japanese and English input based on the prototype show that interactive avatar animations can be displayed in real time and be made available online. This system provides a more natural and interesting means of human communication, and therefore is expected to be used for cross-cultural communication, multiuser online games, and other entertainment applications.

***Keywords*—**personalized avatar; mutiple natural luanguage processing; social backgrounds; anmimation; human computer interaction.

## I. Introduction

AVATARS are graphical representations of real persons in virtual environments [1]. Humanoid avatars especially provide realistic-looking facial expressions, poses, body languages and natural-sounding speech to help people enhance their understanding of the content and intent of a message, and improve significantly human computer interaction (HCI) [2]-[4]. However, as real humans, we have different positions and play different roles in society. Our behaviors are influenced by social backgrounds such as culture, profession, religion, to mention a few [5]. Although humanlike avatars are already available, issues such as where the avatars come from, what kind of social context they have, and what personalities or emotions they harbor, are hardly addressed. This paper proposes a novel method to create humanoid personalized avatar through multiple natural language processing.

Natural human communication by avatars is an ideal dream for scientists across a wide range of disciplines. In order to achieve this target, multidisciplinary efforts are required in the areas of human figure modeling, rendering, animation, natural language processing, speech recognition and synthesis, cognitive science, psychology and linguistics, etc. Although many exciting results [2]-[15] have been presented or applied in recent years, endeavor toward this goal is still a daunting task. The key challenges in this area lie in how to represent personalities and emotions of avatars in terms of social context, how to animate realistic-looking emotional facial expressions, and how to synchronize body or lip movements and facial expressions naturally and realistically. In particular, how to create and animate personalized avatars through natural language input is rarely studied.

Personality is a pattern of behavioral, temperamental, emotional, and mental traits that distinguish people from one another. Personality is a relatively stable tendency that a person carries through a long-term life. On the other hand, emotion is short-lived, influenced by particular events, agents or objects. We assume that each avatar acts associated with distinct personality and emotion as a real human does. This consideration increases reality and believability of avatars. A few toolkits have been implemented with natural language instruction to avatar personalized behaviors.

The virtual human presenter [6] accepts speech text input with embedded commands, which animate the presenter's gestures. A command specifies a motion to coincide with the utterance of a word following the command in the input. One advantage of this system is that it can detect the presence of the corresponding concepts in the raw text stream and automatically insert gesture commands solely on the basis of words used. While the presenter can speak the text together with his actions, it is yet incapable of synchronizing its nonverbal movements with speech at the level of individual words or syllables. And the impact of personality and social context on the presenter remains in future extensions.

Behavior Expression Animation Toolkit (BEAT) [7] is an animation toolkit that automatically extracts actual linguistic and contextual information from text to suggest appropriate hand and arm gestures, facial expressions, and intonation of voice. The mapping from typed text to expressive behaviors relies on a set of rules derived from the state of the art in nonverbal conversational behavior research. Moreover, BEAT allows animators to insert their preferable personalities, motion characteristics, or other particular features in the final animation. BEAT comprises three main processing modules: Language Tagging, Behavior Generation and Behavior Scheduling. The imperfect computational linguistics in the Language Tagging module is the biggest obstacle to BEAT.

Yang et al. [8] propose an interface to extract nonverbal

*Corresponding author. The project is sponsored by the scientific research foundation for the returned overseas Chinese scholars, State Education Ministry, under Grant Q 024131103010068.

Open Science Index, Computer and Information Engineering Vol:3, No:11, 2009 publications.waset.org/13837/pdf

information such as human emotions and intentions embedded in contexts, and to embody them with matched suitable facial expressions and body languages through virtual avatars. But this prototype appears immature, and an intelligent nonverbal information search engine is expected to be implemented in the future.

As stated above, a few pioneer researches have worked toward the personalized or emotional avatars. But there is no mature, easy-to-use and extensible method for creating personalized and compact avatars in virtual worlds to date. Especially, all such efforts do not take multiple natural languages into consideration that are very indispensable in cross-cultural communication. Therefore, we present basic research in this paper leading to the development of methodologies and algorithms for constructing multiple natural language-driven emotive humanoid personalized avatar animations. Our natural language processing algorithms can parse Chinese, Japanese and English at the same time. Not only are these natural languages translated into nonverbal behaviors of avatars, but also some typing errors are tolerated to some extent. Our avatars are model-based using Extendible three-dimensional (X3D) language, and personalities are added automatically to the avatar models according to the background such as culture, profession, age, and so forth. In order to validate our approach, we implement a prototype in Windows XP platform with IE embedded with X3D browser installed, based on browser/server topology. Experiments demonstrate the effectiveness and efficiency of the proposed system.

The rest of this paper is organized as follows. Section II introduces our general framework of the multiple natural language-driven personalized avatar animation synthesis. Section III describes our approaches to the various aspects of the system design. Section IV demonstrates the effectiveness and evaluates the performance of our system with experiments. Finally, section V concludes the paper.

## I. System Overview

This section gives an overview of the proposed system based on a browser/server topology structure, and illustrates the flow chart of animation retrieval.

### A. System Framework

The proposed system uses a browser/server topology structure, as illustrated in Fig. 1, including three main parts: browser, server, and database (DB). The server employing Java Server Pages (JSP) with tomcat, MyEclipse Enterprise Workbench and Java Development Kit (JDK) installed, retrieves and visualizes animations of avatars from the DB. Communication between the server and DB is through Java DataBase Connectivity standard (JDBC), which is an Application Programming Interface (API) when Java invokes DB. DB stores keywords linked with X3D animation files. Once the input natural language text (Chinese, Japanese, English, etc.) matches the keywords in DB, the X3D animation files are retrieved orderly, and the retrieval results are propagated to the user website by JSP. Users input natural language texts in the user interface, and then view retrieval

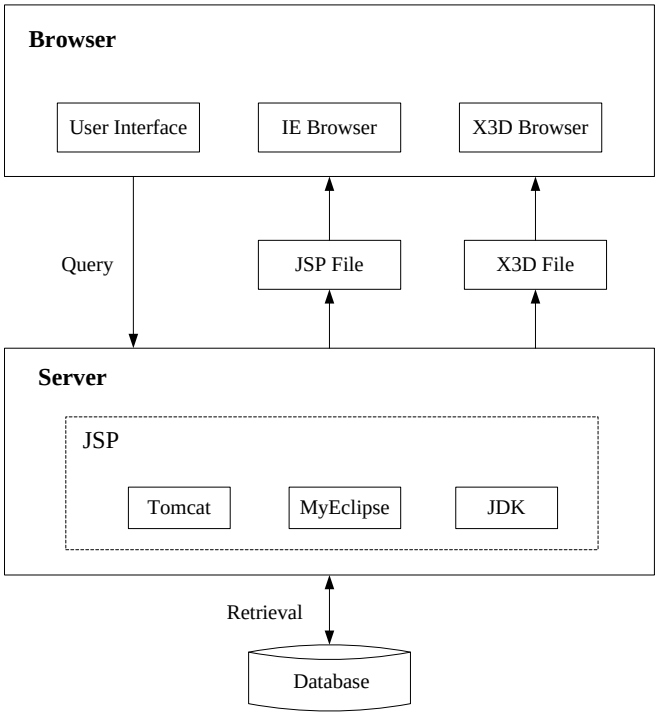



Fig. 1 System framework.

results simply through an Internet Explore (IE) embedded with X3D browser. So this system is independent of the environment of users that makes it very applicable.

### B. Flow Chart of Animation Retrieval

The flow chart of animation retrieval illustrated in Fig. 2 is described as follows.

1) *The server is ready for responding a query from a user.*
2) *A user is connected to the server through the user website.*
3) *A natural language text is input from a user interface.*
4) *The text is parsed to keywords by the natural language processing algorithm.*
5) *Check if a keyword hits in the DB. If yes, retrieve the animation file and then check the remainder of the keywords. If not, end.*

## II. Methodology

This section designs a 3-D humanoid model, which is not only humanlike, but also represents personality and facial emotions based on culture, profession, mood, age, taste, and so forth. And the animation synthesis technique and natural language processing algorithm used in this system are also addressed.

### A. Hierarchical Humanoid Model

Currently, a number of standardization efforts continue to solve different aspects of representing virtual humans in networked virtual environments. One of the most significant

Open Science Index, Computer and Information Engineering Vol:3, No:11, 2009 publications.waset.org/13837/pdf

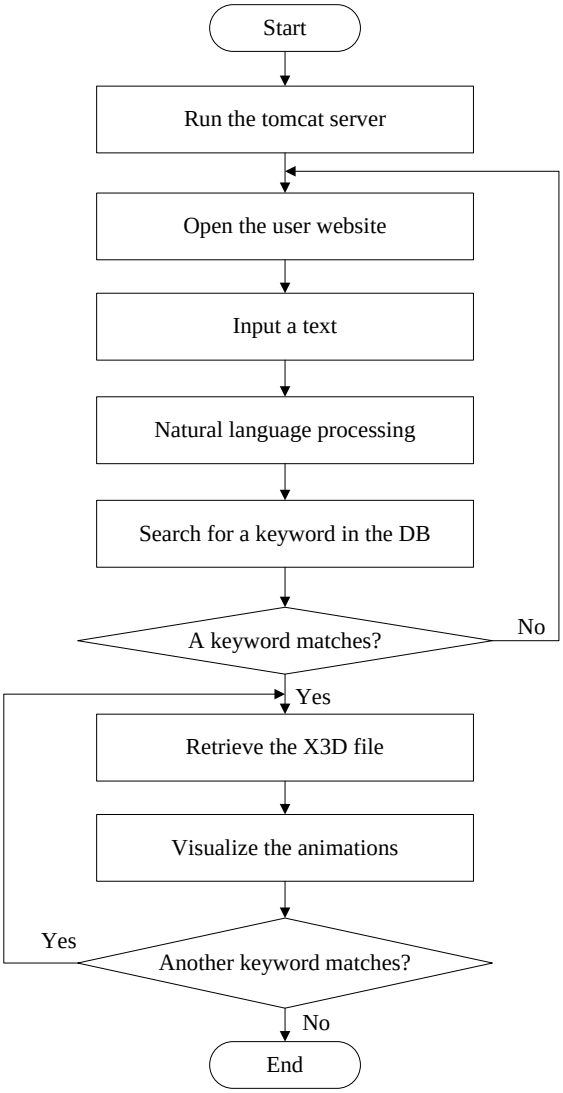



Fig. 2 Flow chart of animation retrieval

among them is Human Animation (H-Anim) specification [16]. Since H-Anim models are easily exchanged between different users or programs, we design the three-dimensional (3-D) articulated humanoid models based on X3D complying with H-Anim standard. This H-Anim figure is formed by assembling H-Anim objects, including Humanoid, Joint, Segment, Site and Displacer. Body parts (segments) are created by X3D-based elements. The 3-D model consists of a number of segments (such as the head, hand, and foot), which are connected to each other by joints (such as elbow, wrist, and ankle) to form a hierarchy. Moreover, DEF and USE defined in X3D-based primitives are employed to define and reuse elements of avatars. Fig. 3 shows two avatars based on this structure.

### *B. Personalized Avatars*

We design the avatars which are not only humanlike, but also represent personalities and facial emotions based on culture, profession, mood, age, taste, and so forth.



Fig. 3 X3D-based hierarchical avatars

Since most of practical applications today involve virtual humans, often crowds, with clothes. In comparison with undress body models, little attention has been devoted to the task of dressing automatically virtual human models [17]. In our study, we design a set of costumes for various avatars from different culture backgrounds and professions. For example, there are 54 ethnic minorities both in China and Vietnam. Fig.4 shows an interface which allows users to dress avatars in their traditional costumes. The appearances of avatars can be changed according to their professions as well. A virtual teacher and a virtual doctor wear different uniforms as shown in Fig. 5.




Fig. IV An interface which allows users to dress avatars in their traditional costumes

Open Science Index, Computer and Information Engineering Vol:3, No:11, 2009 publications.waset.org/13837/pdf

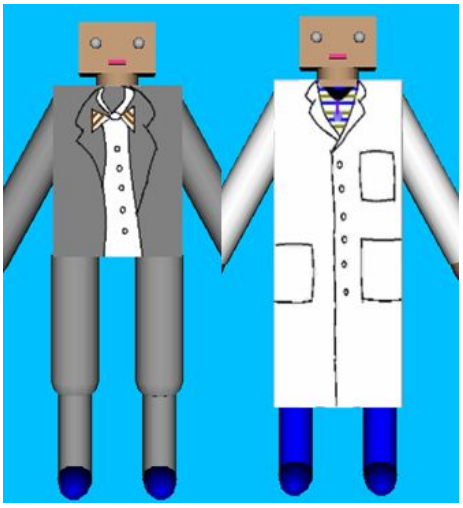


Fig. 5 A virtual teacher and a virtual doctor wearing different uniforms

As for the facial expressions, with some objects embedded in face, we metamorphose the eyes, mouth and eyebrows by extension/flection and rotation functions provided by X3D. Fig.6 shows some facial expressions our avatar plays such as puzzled, amazed, excited, amused, queried etc. The eyes and mouth are designed with some sphere objects, which are metamorphosed to create blinked eyes or opened mouth, as shown in Fig.7. The eyebrow is a mixture of two cylinders half-embedded in the face, and animating the eyebrow contributes a lot to the facial emotions.

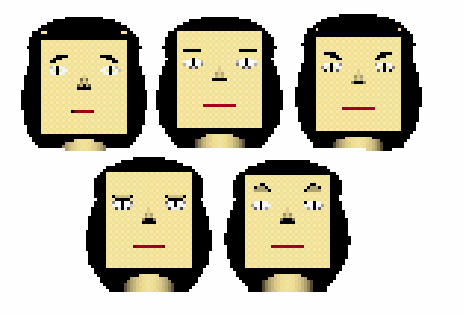


Fig. 6 Facial expressions

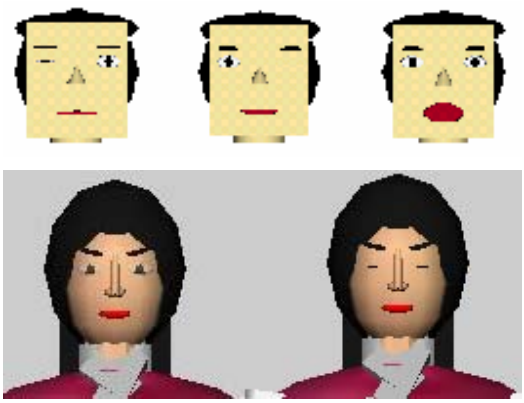


Fig. 7 Animations of eyes and mouth

The age feature is represented by adding wrinkles and warts on the epidermis of avatars. The skin colors can also be changed based on the backgrounds of avatars. Users can select their favorite colors of clothes and shoes according to individual taste. Besides, in order to enhance the believability, we provide suitable dynamic backgrounds with sound by MPEG (Moving Picture Experts Group) files for personalized avatars, as shown in Fig.8.

### *C. Animation Synthesis And Natural Language Processing*

Most of the current communication systems, including various commercial products, are restricted to audio communication and text-based chat capability. Some systems include a means of gestural communication by choosing some icons of predefined gestures or simple behaviors [18]-[22]. But giving avatar run-time instruction from natural language is still relatively unexplored. We aim at synthesizing online real-time avatar animations of body gestures and facial expressions from natural languages. To achieve this, a set of operation and mapping is required, including semantic extraction from the input text, and translation between natural languages and nonverbal languages.

For animation synthesis, besides seamless body gestures, we also create rich facial expressions with appropriate eyes, eyebrows, nose and mouth movements, to target human emotion (such as fear, happiness, despair, etc.) and physiological state (such as hungry, thirsty, etc.). As stated above, we design the skeletal H-Anim avatars with the consideration of simple computation and real-time animation synthesizing. Supposing that the body is in the global coordinate (X, Y, Z), each segment, such as an arm, leg, may have its own local coordinate (X', Y', Z'). Animations can be implemented either in the local coordinates or in the global coordinate. The animations of avatars are created in X3D-Edit by using Timesensor, OrientationInterpolator, and Route.



Fig. 8 A Tibetan avatar with dynamic background

Open Science Index, Computer and Information Engineering Vol:3, No:11, 2009 publications.waset.org/13837/pdf

Our system can be regarded as a system for translating natural languages into 3-D animations. The query text is parsed and the keywords are detected, and then the keywords lead to animations based on the DB. The overall procedure can be described as follows. First, the redundant words such as auxiliary verbs in Japanese are erased from the text automatically through natural language processing algorithm, i.e., just the semantic keywords are extracted. Then the animations are retrieved and displayed according to the order of the keywords. For example, when a Japanese sentence "Watasi (Wa) Oyogi (Ni) Iku" is input, the unnecessary words "Wa", "Ni" are deleted so that the keywords "Watasi", "Oyogi" and "Iku" remain. Then the keywords are translated into the animations orderly.

The following natural language processing algorithm is used for effective indexes.

1) Treat the input message as long as possible and start to search the extracted keywords in the DB.
2) Check if any keyword hits in the DB. If yes, extract its animations and then search the remainder of the original message. If not, end.
3) If the text hits more than one candidate, give priority according to the historical match.

The semantic extraction methodology enables a semantic mapping, ignoring the redundant words or type errors. The above algorithm creates an efficient matching between the input message and the data registered in the DB.

### *D. Database*

We use MySQL as the database which is associated with the MyEclipse in the server. In the DB, each kind of natural language is allotted a repository, e.g., Japanese repository, Chinese repository, and English repository. The DB links with both the natural language keywords and the 3-D animation files. According to the keywords, the arrangement of animations such as which part (left arm, right foot, etc.), what type (rotation, translation, extension, etc.) and play time period of movements is specified. For example, the keywords instruct to rotate or translate an object from one place to another in a period, specify the coordinate values at certain time for certain body part, or give the rotation axis and angle, etc. Therefore, animations are initiated and synthesized based on the run-time instruction from multiple natural languages. The DB cooperates to retrieve animations by the server.

## III. Experiment

This proposed system is developed in Windows XP platform with IE embedded with X3D browser installed. It is an online system to create personalized avatars with proper body movements and facial emotions from multiple natural language input. Fig.9 shows the user interface of the proposed system. The input natural languages include Japanese, Chinese, and English currently. Once a word or sentence is input from a user, the associated animations are displayed below through X3D browser. This system creates matching and mapping mechanisms between natural languages and 3-D animations,

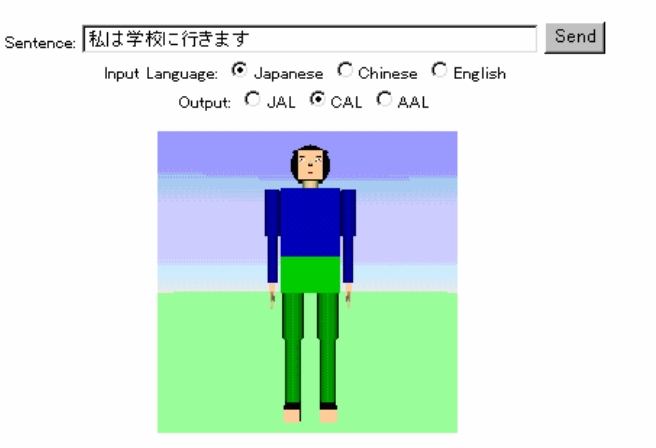



Fig. 10 The user interface

and provides intuitive communication and interaction between the avatar and the novice users without any training needed. Here we illustrate this animation retrieval process with a simple example for this extended abstract.

1) The user input a Chinese sentence "Wo(I) Qu(go) You Yong(swimming)" from a keyboard.
2) This message is sent to the server immediately.
3) The server processes the text and extracts the semantic keywords, i.e., "Wo", "Qu", "Youyong", using natural language processing algorithm.
4) Search for "Wo" in DB and retrieve the associated animation file.
5) The server propagates the retrieval result of "Wo" to the user interface through JSP.
6) Search for "Qu" in DB and retrieve the associated animation file.
7) The server propagates the retrieval result of "Qu" to the user interface through JSP.
8) Search for "Youyong" in DB and retrieve the associated animation file.
9) The server propagates the retrieval result of "Youyong" to the user interface through JSP.

The consecutive animation result based on this sentence is shown in Fig. 10. All the animations are synthesized and the retrieval results are propagated to the user interface immediately. Hence, the user can view the animations corresponding with his input in real time.

The main contributions of this system lie in the following aspects. A 3-D model-based real-time nonverbal communication system with personalized avatars is implemented, which offers more details for natural HCI and provides a platform for 3D model-based real-time animation generation involving two-handed gestures and facial expressions. In addition, effective natural language processing algorithms among multiple languages are proposed. Based on natural language analysis for the Japanese, Chinese and English, we propose a series of processing algorithms. These algorithms can not only delete the redundant words for sentences, but also

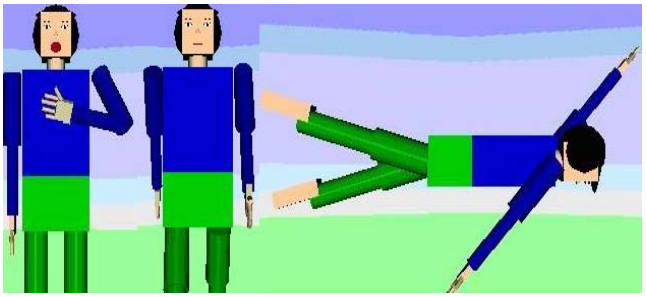


Fig. 11 Consecutive animation result based on a Chinese sentence "Wo Qu Youyong"

tolerate some type errors to a great extent. In addition, matching the sentence as long as possible and arranging the index order of priority in consideration of the record of the past, offer an efficient matching methodology.

## IV. Conlusion

This paper presents a real-time online system to animate humanoid personalized avatars through multiple natural language processing. The 3-D personalized humanoid model design methodology, the animation synthesis technique, and the natural language processing algorithm used in this system are addressed. We have obtained preliminary but promising experiment results as well as built a fully functional system of text-driven emotive personalized avatars based on social backgrounds. Anyone can access this system via the Internet anywhere. Continuous improvements to the system are being made. Although the implemented prototype provides a user-friendly interface with natural languages input, the current version of this system is only for the Japanese, Chinese and English input. Expanding the input into other natural languages could make this system more attractive and practical for the users worldwide. Also, it would be more interesting if we could include other social aspects such as religion for the avatar. This system can be used for cross-cultural communication, multiuser online games, and other entertainment applications. Taking a long view, this technology is expected to provide some contributions to humanoid robot international communication.

## References

[1] T. Capin, L. Pandzic, N. M. Thalmann, and D. Thalmann, Avatars in Networked Virtual Environments, John Wiley & Sons Ltd, England, 1999.

[2] H. Tang, Y. Fu, J. L. Tu, M. Hasegawa-Johnson, and T. S. Huang, "Humanoid audio–visual avatar with emotive text-to-speech synthesis," IEEE Trans. Mul., vol. 10, no. 6, pp. 969–981, October 2008.

[3] Y. Fu, R. Li, T. S. Huang, and M. Danielsen, "Real-time multimodal human–avatar interaction," IEEE Trans. CSVT, vol. 18, no. 4, pp. 467-477, April 2008.

[4] X. L. Yang, D. C. Petriu, T. E. Whalen, and E. M. Petriu, "Hierarchical animation control of avatars in 3-D virtual environments," IEEE Trans. Instrumentation and Measurement, vol. 54, no. 3, pp. 1333-1341, June 2005.

[5] J. Hou, Y. Koine, and Y. Aoki, "Ontology-based social avatar language," Journal of Signal Processing, vol. 7, no. 5, pp. 393–403, Septemper 2003.

[6] T. Noma, L. W. Zhao, and N. Badler, "Design of a virtual human presenter," IEEE Comput. Graph. Appl., vol. 20, no. 4, pp. 79–85, July 2000.

[7] J. Cassell, H. Vilhjalmsson, and T. Bickmore, "BEAT: the behavior expression animation toolkit," in Proc. 28th Int. Conf. Computer Graphics and Interactive Techniques, Los Angeles, USA, Aug. 2001, pp. 477–486.

[8] Z. X. Yang, L. Li, and D. Zhang, "Embodiment of text based on cirtual robotic avatar," in Proc. IEEE Int. Conf. Robotics and Biomimetics, Sanya, China, Dec. 2007, pp. 1285 – 1289.

[9] W. Steptoe and A. Steed, "High-fidelity avatar eye-representation," in Proc. IEEE Virtual Reality Conf., Nevada, USA, Mar. 2008, pp. 111-114.

[10] Z. Sakr and C. Sudama, "A curvilinear avatar with avatar collision detection scheme in collaborative virtual environments," in Proc. IEEE Instrumentation and Measurement Technology Conf., Victoria Vancouver Island, Canada, May 2008, pp. 1027 – 1030.

[11] Y. Ishii and T. Watanabe, "An embodied avatar mediated communication system with VirtualActor for human interaction analysis," in Proc. 16th IEEE Int. Conf. Robot and Human Interactive Communication, Jeju, Korea, Aug. 2007, pp. 37-42.

[12] S. Piotr and K. Bozena, "Personalized avatar animation for virtual reality," in Proc. 1st Int. Conf. Information Technology, Gdansk, Poland, May 2008, pp. 1-4.

[13] S. Rusdorf, G. Brunnett, M. Lorenz, and T. Winkler, "Real-time interaction with a humanoid avatar in an immersive table tennis simulation," IEEE Trans. Visualization and Computer Graphics, vol. 13, no. 1, pp. 15-25, January-February 2007.

[14] C. Lee, H. Lee, and K. Oh, "Real-time image-based 3D avatar for immersive game." in Proc. 7th ACM SIGGRAPH Int. Conf. Virtual-reality Continuum and Its Applications in Industry, Singapore, Dec. 2008, pp. 1-2.

[15] N. Ahmed, E. Aguiar, C. Theobalt, M. Magnor, and H. P. Seidel, "Automatic generation of personalized human avatars from multi-view video." in Proc. the ACM Symposium Virtual Reality Software and Technology, Monterey, USA, Nov. 2005, pp. 257-260.

[16] The Humanoid Animation Group's. [Online]. Available: http://www.hanim.org/.

[17] N. M. Thalmann, H. Seo, and F. Cordier, "Automatic modeling of animatable virtual humans-a survey," in Proc. 4th Int. Conf. 3-D Digital Imaging and Modeling, Banff, Canada, Oct. 2003, pp. 2-10.

[18] I. J. Pelczer, F. Cabiedes, and F. Gamboa, "Emotions and interactive agents," in Proc. IEEE Int. Conf. VECIMS, La Coruña, Spain, Jul. 2006, pp.184-187,.

[19] Globe Warp's web site [Online]. Available: http://chat.globewarp.or.jp/index.html

[20] T. Wang, X. Li, and J. Shi, "An avatar-based approach to 3D user interface design for children," in Proc. IEEE Symposium 3D User Interfaces, North Carolina, USA, Mar. 2007, pp. 155-161,.

[21] Sony's 3-D Chat web site [Online]. Available: http://www.so-net.ne.jp/paw

[22] Y. J. Oh, K. H. Park, and Z. Bien, "Body motion editor for sign language avatar," in Proc. Int. Conf. Control, Automation and Systems, Seoul, Korea, Oct. 2007, pp. 1752-1757.

Open Science Index, Computer and Information Engineering Vol:3, No:11, 2009 publications.waset.org/13837/pdf