Journal of Korea Multimedia Society Vol. 13, No. 6, JUNE 2010(pp. 924-932)

A Multi-Modal Complex Motion Authoring Tool for Creating Robot Contents

Kwang-Ho Seok†, Yoon Sang Kim††

ABSTRACT

This paper proposes a multi-modal complex motion authoring tool for creating robot contents. The proposed tool is user-friendly and allows general users without much knowledge about robots, including children, women and the elderly, to easily edit and modify robot contents. Furthermore, the tool uses multi-modal data including graphic motion, voice and music to simulate user-created robot contents in the 3D virtual environment. This allows the user to not only view the authoring process in real time but also transmit the final authored contents to control the robot. The validity of the proposed tool was examined based on simulations using the authored multi-modal complex motion robot contents as well as experiments of actual robot motions.

Key words: Robot Content, Authoring Tool, Multi-Modal Complex Motion

1. INTRODUCTION

With rapid developments in today's robot technology, service robots have been used in a wide range of fields. However, robots must become more affordable in terms of price and easier to use for general users in order to expand the scope of applications of intelligent robots in our everyday life. Based on these requirements, researches and developments have been taking place in countries with advanced robot technologies such as the U.S. and Japan regarding various robot software products (authoring tools in this study) that allow users to easily create a variety of robot commands (services). Major software products include ERSP™ from Evolution Robotics [1], RoboStudio™ from NEC [2], UML-RT from OMG(Object Management Group) [3], 3D Simulation Tool from CyberRobotics [4] and Microsoft Robotics Studio(MSRS) [5]. Evolution Robotics developed a mobile robot software development tool called ERSP for vision-based mobile robot control programming, which had required a high level of technical skills. CyberRobotics is applying its robot software development tool Webots to diverse types of robots, and Microsoft entered the robot market with MSRS 1.0, a robot software development tool for Windows. In Korea, a simulator software package based on a flowchart-type graphic language was recently developed to facilitate the research and development of elemental and control technologies for intelligent robots in the 3D virtual environment [6]. The common aim of these software solutions is to improve the user accessibility of robot software development tools to expand the robot market.

Robot motion authoring has been studied with character animation in virtual environments, and thus numerous studies have been applied into the robot motion authoring. In [7], a *motion patch*, which informs what motions are available for ani-

---

※ Corresponding Author : Yoon Sang Kim, Address: (330-708) 1800 Chungjeollo (307 Gajeon-ri) Byeongcheon-myeon Cheonan-city Chungnam Province, Korea, TEL: 041)560-1496, FAX: 041)560-1462, E-mail: yoonsang@kut.ac.kr

Receipt date : May 26, 2009, Revision date : Mar. 18, 2010
Approval date : May 13, 2010

† Department of Information-Media Engineering, Korea University of Technology and Education
†† School of Computer and Science Engineering, Korea University of Technology and Education

mated characters within a block, is presented as a new technique for allowing animated characters to navigate through a virtual environment. An automatic generation of robot motions for effective human-robot interaction is presented in [8], but it is still inconvenient or very limited because the motions are generated by samples thus have the best fit on special cases. In [9], a motion generating system for humanoid robots is presented based on the key-frame animation. Unlike dynamic-based method which motion generation for humanoids has studied mainly, the proposed system generates a robot motion by a sequence of given several postures, and all the motions of *Tai Chi Chuan* are produced by the system.

However, user-intuitive or user-friendly robot command authoring tools that focus on facilitating general users are still rare. The lack of user-intuitive or user-friendly tools is likely to create a barrier for providing robot services (commands/motions) that correspond with the preferences and demands of general users including children, the elderly and housewives, who are expected to be the major clients in the service robot industry.

This paper proposes a multi-modal complex motion authoring tool capable of creating robot contents. The proposed tool is user-friendly and allows general users without much knowledge about robots, including children, women and the elderly, to easily edit and modify robot contents. More specifically, the multi-modal complex motion authoring tool proposed in this paper creates basic motions, links them to author various robot movements and provides a simulation environment to test the authoring outcome. The proposed authoring tool also uses multi-modal data such as graphic motions, voice and music to simulate user-created robot contents in a 3D virtual environment so that the user can view the authoring process in real time and transmit the authored robot contents to control the robot. The validity of the proposed tool was examined based on simulations using the authored multi-modal complex motion robot contents as well as experiments of actual robot motions.

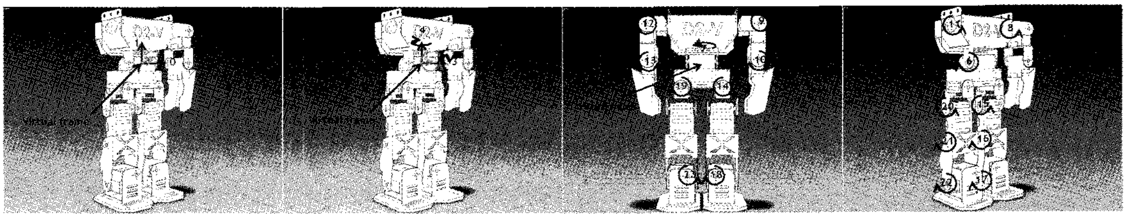
This paper is organized as follows. The proposed multi-modal complex authoring tool for creating robot contents is reviewed in Section II, which includes the media sequencer and motion editor. The experimental results and discussion are presented in Section III, followed by the conclusion in Section IV.

2. MULTI-MODAL COMPLEX MOTION AUTHORING TOOL

The multi-modal complex motion authoring tool implemented by this study divides authoring functions into high-level functions performed by the media sequencer (MS) and low-level functions performed by the motion editor (ME). Each level carries out independent roles. MS aligns pre-authored media clip files on the timeline to create and edit intended motions, and ME assigns the start and end times of movements to author robot motions. The authored outcome is saved as a motion file and played back in MS, and the motion file is implemented with an independent format according to robot characteristics. Media clips aligned on the MS timeline consist of robot voice implemented with TTS(text-to-speech) synthesis, general sound sources used for background audio (*.wav, etc.) and media files that designate robot behavior (bodily motions and facial expressions; behavior is restricted to motions in this study). In addition, motion editing and motion generation are provided to render smooth transition between two different robot motions for graphic representation.

2.1 Robot Structure

For this study, a humanoid with 24 degrees of freedom (DOF) was used as the object of robot contents authoring. The position of the robot constantly changes according to robot motions on the absolute coordinate. Accordingly, a virtual frame



(a) (b) (c) (d)

Fig. 1. Translation and rotation of the humanoid used for robot contents authoring: (a) Translation in X, Y and Z directions of the virtual frame; (b) Rotation in X, Y and Z directions of the virtual frame; (c) Virtual actuators (d) Virtual actuators.

was established above robot's waistline to calculate joint motions on the absolute coordinate according to movements, and each actuator was assigned with a unique ID to represent robot's motions. IDs 0~2 and 3~5 indicate translation and rotation, respectively, in X, Y and Z directions of the virtual frame (Figure 1 (a) and (b)). In addition, each joint of robot has its own virtual actuator to represent the rotation of the joint (Figure 1 (c)).

2.2 Media Sequencer

MS-based authoring involves aligning media files on the timeline. As shown in Figure 2, MS aligns TTS, sounds and motions on the timeline to author intended complex movements and was designed to perform independently from the lower functions of ME. The authoring process consists of assigning media types to the track layer and adding new media files according to the assigned type.

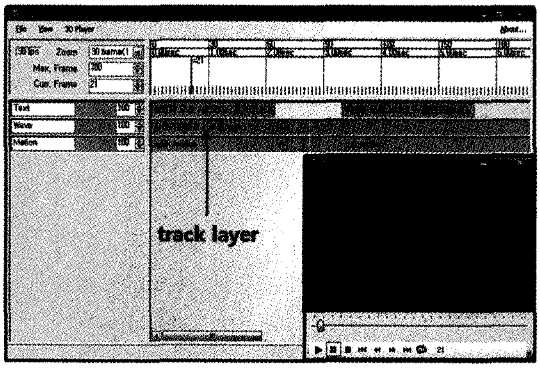


Fig. 2. Media Sequencer.

Different types of media files provide following functions:

(1) Voice Authoring using TTS(Text-Speech Synthesis): User's text input in the MS track is synthesized into a robot voice, providing a talking capability to robot contents. There is only one text string media track, and thus text strings must not overlap in the timeline.

(2) Sound Authoring: User's sound input in the MS track is converted to be used as the background audio or sound effect of the multi-modal content. General sound files can be used for sound authoring, although compatible sound file formats may vary according to the robot hardware. Up to n sound tracks can be mixed on the timeline for simultaneous output. The sound files may be authored by different commercial programs.

(3) Motion Authoring: User's motion input in the MS track is converted to be used as the robot motion of the authored multi-modal content. The motions used for motion authoring are authored with robot's joint and position from ME's lower functions, and are combined with TTS and sounds to create multi-modal robot contents. As with sound files, multiple motion files can overlap on the timeline for playback.

Accordingly, MS allows the user to combine TTS-synthesized voice and sounds with robot motions to provide new types of multi-modal robot contents that include motions, voice and sounds (music). The final result authored by MS is in the form of an event index file that contains the start and end times for each media file. The result also contains processing parameters including sound file volume, motion blending and transition. In other words, only the combinations of the aligned media files are examined in MS, and each media file is stored with separate authoring programs. The final authored files from MS only contain the index of each media clip file.

2.3 Motion Editor

ME provides the functions for creating robot motions. In order to allow users to easily author robot motions, the ME developed for this study, shown in Figure 3, converts the variation of each joint value with the progress of time into a graphic form to simplify the process of creating and modifying robot's motion data. The ME allows authoring of robot motions with an approach similar to those of commercial modeling and animation authoring tools such as 3D Studio Max™, Maya™ and Flash™ so that users familiar with commercial tools as well as those with little robot knowledge including children and the elderly to easily create robot motions. Whereas 2D object-oriented drawing editor such as Microsoft PowerPoint provides easy and intuitive direct manipulation, translation and rotation must be adjusted in three directions (X, Y and Z) for a single joint in 3D, making it difficult to control with mouse or keyboard. Accordingly, most 3D editors provide indirect manipulation using a slide bar as default, and direct manipulation is offered as an additional feature. Based on this general trend, the authoring tool developed for this paper was designed to support both direct and indirect manipulation. A joint value can be typed into the text input window next to the name of the joint name (Figure 3, A). The indirect manipulation approach offered by the tool involves dragging an apex of the graph while adjusting the slide bar (Figure 3, B), and direct manipulation requires clicking a joint of the robot in the 3D rendering window and moving it to a desired location

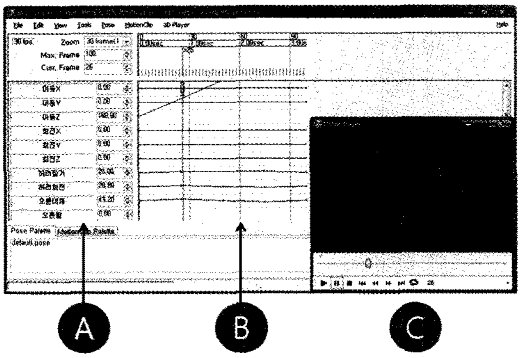
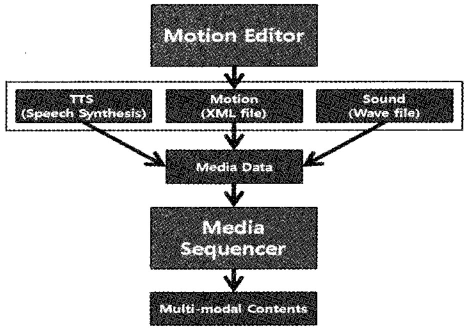


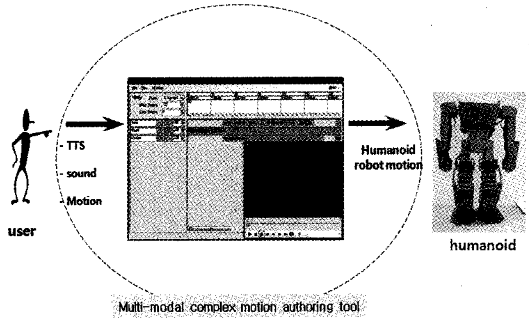
Fig. 3. Motion Editor.

(Figure 3, C). In addition, ME's motion authoring supports both pose and joint-based approaches. Pose-based robot motion authoring creates robot's pose by adjusting joint values according to time, which is an easy way appropriate for beginners. However, it is difficult to link various poses and render natural motions to author high-quality motion output. In comparison, joint-based robot authoring creates robot motions by adjusting joint values required for intended poses. Although this approach can create natural motions with minimum and optimal key frames, skilled manipulation is required to obtain a desired pose by varying joint values.

In the case of a motion file containing n media tracks to be played back with multiple motions at a single point in time (multiple motions being played back simultaneously), there needs to be a resolution to deal with the different values assigned to a single joint by multiple motions. The ME implemented for this study adopts the simple method of using the average of multiple joint values. However, even when two motions are played back simultaneously, applying the average value becomes pointless if one is facial expression and the other is associated with bodily motion. In order to distinguish such a circumstance, it is necessary to identify that a specific region of a specific joint is significant rather than every value of every motion. Therefore, ME must be able to distinguish between joints with and without values when assigning a particular pose. The issue was resolved in this study by assigning a motion value to a specific region of a specific rather than every motion of the robot. ME designates joints with motions in their poses as "key joints", and assigns each key joint with a motion weight factor that indicates the degree of importance of the joint value to link two motions seamlessly. For example, if we assume that one motion is swinging arms back and forth and the other is walking with legs, a new natural motion of swinging arms while walking can be created by applying motion weight factors. However, the weight factor is difficult to apply when there is another key joint representing a different motion at the same time with an identical or similar weight factor. This conflict was resolved with simple motion mixing by applying different ratios between two motions according to time. The joint values were mixed by gradually increasing the ratio of the start motion from 0 to 1 and decreasing that of the end motion from 1 to 0 to yield a seamless transition between the two motions.



(a) Authoring sequence based on the proposed system



(b) Overall blockdiagram of the proposed authoring tool

Fig. 4. Multi-Modal Complex Motion Authoring Tool System.

3. EXPERIMENTAL RESULTS AND DISCUSSIONS

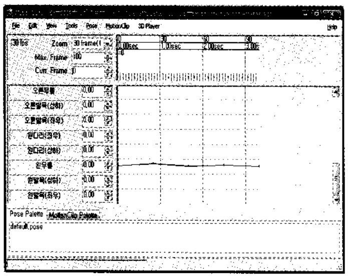
3.1 Multi-Modal Complex Motion Authoring Tool System

The multi-modal complex motion authoring tool system proposed in this paper was implemented on a personal computer with 1GByte memory, 1.80 GHz AMD Athlon™ 64-bit processor CPU and ATI Radeon X1600 graphic card that supports DirectX. Other software packages used for the system include Microsoft DirectX and Microsoft. Net Framework 2.0 as well as Microsoft Speech SDK 5.1 [10], NeoSpeech Korean Yumi16 Voice [11] and TextAloud TTS Interface for providing voice recognition technology to the multi-modal authoring tool. As well, Bluetooth 2.0 was used to transmit(distribute) the authored contents to the robot. Figure 4 illustrates the overall authoring process of the multi-modal complex motion authoring tool system implemented for this study.

3.2 Robot Contents Authoring and Its Application

3.2.1 Experiment 1 : Basic Motion Authoring

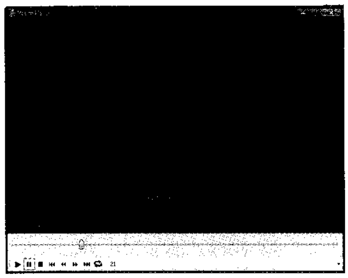
The purpose of Experiment 1 was to confirm whether robot motions authored with ME are played back properly by the Motion Player as well as in an actual robot. Joint-based authoring was used to create robot motions, and the user adjusted joint values to implement continuous movements. Although the user can use codes as in Figure 5(b), ME provides an intuitive approach of creating and authoring desired robot movements by adjusting each joint value as shown in Figure 5(a). The data for expressing robot motions are defined with XML files, which only record key frame rotation values of each virtual actuator and not those of stationary



(a)

```
<?xml version="1.0" encoding="euc-kr" ?>
<motion_clip start_frame="0" end_frame="100"
        frame_rate="30">
    <key_track name="back bending" id="6">
        <key frame="0" val="0.000000" />
        <key frame="25" val="20.000000" />
        <key frame="75" val="20.000000" />
        <key frame="100" val="0.000000" />
    </key_track>
    <key_track name="left knee" id="21">
        <key frame="0" val="0.000000" />
        <key frame="25" val="60.00000" />
        <key frame="50" val="0.000000" />
        <key frame="75" val="30.00000" />
        <key frame="100" val="0.000000" />
    </key_track>
</motion_clip>
```

(b)

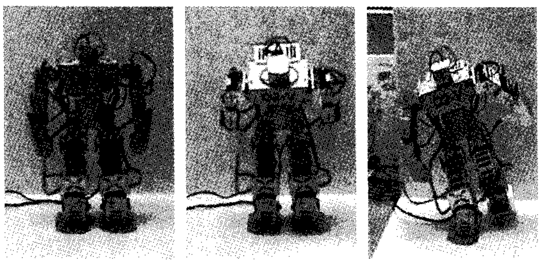


(c)

Fig. 5. Result of basic motion authoring of Experiment 1: (a) Motion authoring using ME, (b) Pseudo code for bending the back and the left knee, (c) Screen capture of motion displayed in Motion Player.

joints. Figure 5(b) displays the XML pseudo code for bending the back and bending the left knee. The motion specified in the pseudo is carried out from Frame 0 to Frame 100 and a frame rate was set at 30 fps. Motion values were assigned to the key frames of ID 6 and 21 that correspond to bending the back and the left knee. Figure 5(c) shows the result of executing the pseudo code of Figure 5(b) in ME Motion Player.

In addition to testing on the Motion Player, the motions authored by ME were applied to an actual robot for the experiment (Figure 6). The motion data was transmitted to the robot using Bluetooth and it was examined whether the contents containing TTS, sound and motion produced by the multi-modal complex motion authoring tool controlled the robot according to user's intention.



(a) (b) (c)

Fig. 6. Robot moving according to motion data authored with ME transmitted using Bluetooth : (a) Start (b) Back bending (c) Left knee bending.

Based on the approach explained above, a total of nine basic motions (elbow rotation motion, shoulder rotation motion, back bending motion, ankle rotation motion, vertical leg rotation motion, kick motion, walk motion, arm rotation motion and waist rotation motion, Figure 7) were created and checked for this study [12,13].

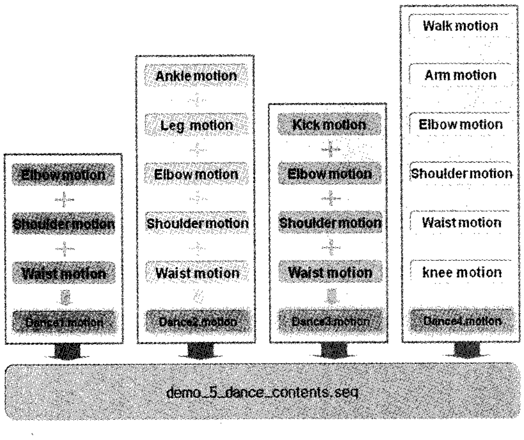


Fig. 7. Example of complex motion authoring using basic motions.

3.2.2 Experiment 2 : Multi-Modal Complex Motion Contents Authoring

Experiment 2 involves examining the possibility of authoring multi-modal complex motion robot contents using the proposed tool. Background audio consisting of sound and TTS voice was added to the motions newly created with various combinations of basic motions and simulations were performed to check actual robot movements, in order to verify whether general users with little knowledge about robots including children, women and the elderly can easily author robot contents.

A dance motion was implemented as shown in Figure 7 based on four motions newly created by combining the nine basic motions produced for robot contents authoring. Figure 8 shows the authoring process of robot contents containing sound and TTS voice created with the multi-modal complex motion authoring tool proposed in this paper. Figure 9 displays screen capture of the simulation result from the proposed authoring tool (multi-modal complex motion robot contents that consist of new motions created by the combinations of basic motions and the background audio containing sound and TTS voice) as well as actual robot movements controlled by the transmitted data. The experimental results confirmed that the multi-modal complex motion authoring tool system proposed by this study provides general users with an intuitive, fun and easy way to create basic motions, which thus a wide range of multi-modal complex motions can also be authored.

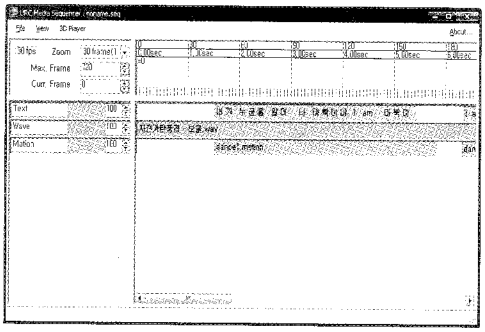


Fig. 8. Multi-modal contents authoring using MS.

The comparison between the conventional methods and the proposed system is summarized as the following table.

4. CONCLUSION

This paper proposed a multi-modal complex motion authoring tool for creating robot contents, which is composed of two parts: the ME (motion editor) for creating robot motion and the MS (media sequencer) for editing various media files such as voice, music and (graphic) motion. Two experiments were explored on the proposed authoring tool. From the result of experiment 1, the functionality for basic motion authoring was checked and a total of nine basic motions were created. The possibility of authoring multi-modal complex motion robot contents were examined by adding various media files (such as sound and voice) into the motions in experiment 2. The effectiveness of the proposed tool was confirmed based on the simulations using the authored multi-modal

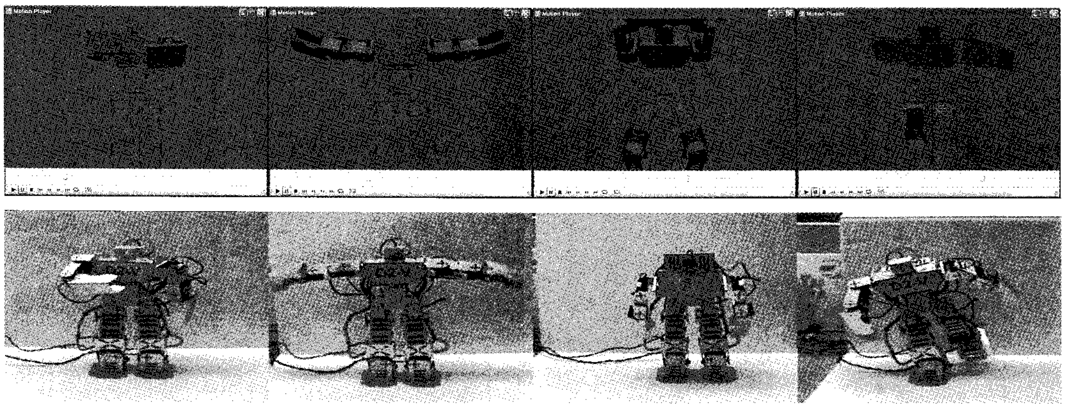


Fig. 9. Simulations of robot contents authored with the proposed multi-modal complex motion authoring tool and their corresponding actual robot movements.

Table 1. Comparison between the conventional methods and the proposed system

Comparative indices / Each method	main role	skill level for user's accessibility	graphic simulation	transfer to actual robot	applicable robot
[1]	software developer	High	Yes	Yes	Wheeled
[2]	authoring	Fair	Yes	Yes	Specific
[4]	software developer	High	Yes	Yes	Wheeled, Humanoid
[5]	software developer	Fair	Yes	Yes	Wheeled
[6]	authoring	Low	Yes	Yes	Wheeled
Proposed	authoring	Low	Yes	Yes	Humanoid

complex robot contents as well as the experimental results of the robot contents including motions transmitted to an actual robot.

REFERENCES

[1] http://www.evolution.com/(Evolution ERSP™)

[2] http://www.necst.co.jp/english/company/laboratory. htm#01 (NEC RoboStudio™)

[3] http://www.omg.org (Object Management Group. UML-RT)

[4] http://www.cyberbotics.com/ (CyberRobotics Webots™)

[5] http://www.microsoft.com/korea/robotics/studio.mspx

[6] S.Y. Nam, H.Y. Lee, S.J. Kim, Y.C. Kang, and K.E. Kim, "A study on design and implementation of the intelligent robot simulator which is connected to an URC system," Vol. 44, No.4, *IE*, pp. 157-164, 2007.

[7] K.H. Lee, M.G. Choi, and J.H. Lee, "Motion Patches: Building Blocks for Virtual Environments Annotated with Motion Data," *ACM Transaction on Graphics (SIGGRAPH 2006)*, Vol.25, No.3, pp. 898-906, July 2006.

[8] K.W. Sok, M. Kim, and J.H. Lee, "Simulating Biped Behaviors form Human Motion Data," *ACM Transactions on Graphics (SIGGRAPH 2007)*, Vol.26, No.3, Article 107, July 2007.

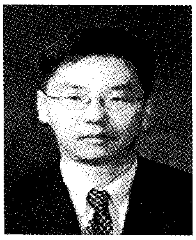
[9] K. Noritake, "A Motion Generation System for Humanoid Robots - Tai Chi Motion -," International Symposium on Micromechatronics and Human Science, pp. 265-269, 2003.

[10] http://www.microsoft.com/downloads/ (Microsoft Speech SDK 5.1)

[11] http://www.nextup.com/ (NeoSpeech Korean Yumi16 Voice)

[12] K.H. Seok and Y.S. Kim, "A Multi-Modal Complex Motion Authoring Tool for Creating Robot Contents," Proceedings of the 39th International Symposium on Robotics (ISR), pp. 593-597, October 2008.

[13] S.Y. Kim, J.O. Kim, K.H. Seok, B.S. Seo, J.B. Kim, and Y.S. Kim, "Development of the Multi-Modal Complex Motion Authoring Tool for URC Robot," Proceedings of the 2nd KRS Korea Robotics Society Annual Conference, pp. 474-479, 2007.



Yoon Sang Kim

the B.S., M.S., and Ph.D. degrees in electrical engineering, from Sungkyunkwan University, Korea, in 1993, 1995, and 1999, respectively. He was a Member of the Post-Doctoral Research Staff of the Humanoid Robotics Research Center, Korea Institute of Science and Technology (KIST), Seoul, Korea. He was also a Faculty Research Associate in the Department of Electrical Engineering, University of Washington, Seattle. He was a Member of the Principal Research Staff, Samsung Advanced Institute of Technology (SAIT), Suwon, Korea. Since March 2005, he has been an Associate Professor at the School of Computer Science and Engineering, Korea University of Technology Education (KUT), Cheonan, Korea. His current research interests include immersive virtual reality, haptic technology, and robotics.

Kwang-Ho Seok

the B.S. degree in internet- media engineering, from Korea University of Technology and Education (KUT), Korea, in 2007 and the M.S. degree in Information media engineering, from Korea University of Technology and Education (KUT), Korea, in 2009. Since 2009 to now, he has been Ph.D. student in Humane Interaction Lab (HILab), Korea University of Technology and Education (KUT), Cheonan, Korea. His current research interests include immersive virtual reality, human robot interaction, and power system.