Multiple-degree-of-freedom motion data transmission using audio signal and its application in a presentation robot system

S. OWA, T. IZUMI, S. IWAKI (Hiroshima City University), and A. NAKAYAMA (NTT)

Abstract— **Employing an audio signal as a data carrier, we study a wired data transmission method for a multi-degree-of-freedom robot. In this paper, we propose a frequency division multiplexed amplitude modulation method and its demodulation parameter design method to minimize the processing time. To verify the proposed method, an experimental system is implemented on a single PC with a 3DOF robot. As a practical application, a presentation robot system synchronized with a slide show in PowerPoint is introduced.**

I. INTRODUCTION

Recently, the use of robots has been tested for the automated and effective presentation of various types of information, such as product introductions and in-store guidance [1-4]. These presentation robots convey intentions and feelings to humans by performing various movements. It is rare that these motions are reproduced alone, and they are often accompanied by sound content such as voice, music, or sound effects. In many cases, the action of the presentation robot and the sound have an inseparable relationship and it is therefore essential that both be accurately reproduced in synchronization. However, accurately synchronized playback is frequently a difficult problem with regards to non-real-time processing and multitask environments. In particular, an unstable misalignment in the playback of both frequently occurs due to other tasks being processed. This problem, which also sometimes occurs when video and sound are played back, causes a sense of discomfort in the audience and is a cause of robots losing their lifelike feel.

Against this background, the concept of using sound or the sound signal itself as a transmission medium for motion data (MD) has recently gathered attention as a method for the simultaneous playback of sound and motion. In pioneering research by Nakayama et al., robot movement was treated as motion media, and from the perspective of the distribution of this content, the significance of the MD transmission by sound was considered [5]. For a single degree of freedom, they proposed an MD transmission method using dual-tone multi-frequency and frequency modulation signals as concrete techniques, and applied this to the control of a one-degree-of-freedom robot. However, a multichannel transmission method for a multi-degree-of-freedom robot was not considered. Meanwhile, Gerasimov [6] and Lopes [7] comprehensively discussed various methods of data transmission by sound and the effects on human hearing, for applications other than robots. However, there was no consideration of the application to robots, and so the discussion of a method suited to robot control was lacking in those studies also. As above, design theory relating to multiple degrees of freedom, such as what type of multiplexing technique is desirable, or to what extent multiple degrees of freedom are possible, has not been examined in relation to MD transmission using sound. Accordingly, this paper first considers methods for multiple degrees of freedom and proposes a multi-degree-of-freedom robot MD transmission method based on frequency division multiplexed amplitude modulation using an audio signal as a carrier wave. In addition, a robot system that automatically performs a presentation linked with Microsoft PowerPoint, the current de facto standard presentation software, is proposed as an application example.

This paper shows the merits in using sound as a robot data transmission method and considers two methods (encoding/decoding method and modulation/demodulation method) for doing so. A frequency-division multiplexed amplitude modulation method is then proposed, and a demodulation processing parameter determination method that gives consideration to minimization of the demodulation processing time is presented. Next, the proposed method is implemented in an automatic presentation robot system that utilizes the sound playback function of Microsoft office PowerPoint, and the validity of the method is verified by means of a performance test with a three-degree-of-freedom robot. Finally, the potential of this method is demonstrated through the presentation of various robot content examples that use the proposed method.

II. ADVANTAGES OF USING SOUND AS A TRANSMISSION MEDIUM

In usual transmission methods, the transmission medium of electromagnetic waves, electricity, or light transports a text, sound, image, or video signal. Conversely, in this research, sound or the audio signal itself transports the MD (motion data). The advantages to using sound as a transmission medium for MD in this way, including the aforementioned synchronicity, are given below.

1) Affinity with existing sound resources: The sound infrastructure, devices, and software that already exist in abundance throughout the world can be used. For example, the sound infrastructure includes the Internet, broadcasts, and telephone lines; devices include CD players and computer terminals; and software includes sound-related application programming interfaces (APIs) and plug-ins, and terminal software. By making use of these, the opportunity for robot content to be circulated, distributed, and accumulated throughout the world is increased and a diverse style of communication can be constructed.

2) Synchronization with audio media: The MD is embedded in the audio signal itself and thus the synchronization of the MD with audio content is straightforward. This is an essential property for content in which sound and movement are required to be in perfect alignment, such as lip syncing or dance. It also removes the necessity of considering troublesome synchronization between media during content playback.

3) Simplicity of physical interface: Analog sound terminals and speakers are the most commonplace interface in the world. Although the wave of digitalization shows no sign of stopping, it is anticipated that for the foreseeable future, an analog interface will continue to play the active role in regard to speakers, the final exit point of sound. If this general-purpose analog interface can be utilized as a transmission route for robot data, the venues where robot applications can be implemented will greatly increase.

4) Ease of application creation: If the environment has an arrangement that produces sound, a linking application for sound and movement can be easily produced. By simply exchanging the processing section that generates sound or the sound file for the application (ringtone, application start-up sound, etc.) for an existing application that produces sound, it is possible to modify it into application software by which the text, image, or picture of a PC screen is linked with a robot. In addition, by similarly reusing the sound file, the reuse of motion content becomes possible.

In the above way, an MD transmission method using sound is likely to be highly applicable as a sound/motion playback technology for various robots and can be said to be desirable for the presentation robots that are the subject of this paper.

III. Multi-degree-of-freedom MD transmission method by sound

In this section, two existing methods (encoding/decoding and modulation/demodulation method) are taken under consideration and the more appropriate multi-degree-of-freedom MD transmission method using sound is selected.

A. Coding/decoding method:

In this method, a "symbol" or "command" is sent; specifically, the MD is encoded using the motion command registered in the code directory, and then decoded at the receiver after having been transmitted using sound as a carrier wave (Fig. 1(a)). As this is command-based MD transmission, operation is simple for the user and it is highly universal as it can be applied to any robots that support the code directory. However, as the decoding process is generally complex and requires time, the synchronicity with the sound data deteriorates easily. In addition, intelligence for interpreting the command is necessary inside the robot itself. Furthermore, if the command that the transmitter sends is not registered in the receiver, an enquiry cannot be made and so synchronization of the code directory between the transmitter and receiver is necessary.

B. Modulation/demodulation method:

In this method, a "signal" is sent; specifically, time series data for motion are modulated/demodulated with respect to the sound that is the carrier wave without encoding of the target values of the joint angle parameter or fingertip position for the robot. The advantages of this method are that processing is simple and so synchronicity can be high, and that implementation of the processor for the demodulation side is possible with only cheap analog circuits, which in turn should lead to increased availability of robots in a variety of fields. Conversely, the disadvantage of this method is that it is necessary to send a command for each degree of freedom for every operation, which means that changes in robot kinematics cannot be supported.

This report places the synchronized playback between the sound and MD as the most important item and will only discuss the modulation/demodulation method hereafter.

IV. Proposed technique based on the modulation/demodulation method

A. Conditions sought in the motion data transmission method.

Here, the conventional method of radio broadcasting is discussed as a representative example of sound signal transmission by means of a typical electromagnetic wave. Through comparisons with this example, the principal requirements sought in robot MD transmission are arranged as follows.

1) Narrow frequency band of carrier wave: The carrier frequency region for AM broadcasting in Japan is from approximately 150 KHz to 30 MHz, whereas very-high frequency (87.5-108 MHz) is used for FM broadcasting, which has good sound quality. In comparison to these, the proposed transmission band is extremely narrow at approximately 20 Hz - 20 KHz in order to match the specifications of the input, output, and processing apparatus for sound. The way in which to effectively allocate the MD to this narrow frequency band is a technical challenge.

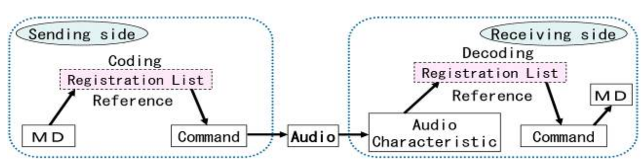


Fig. 1(a) Coding/Decoding method

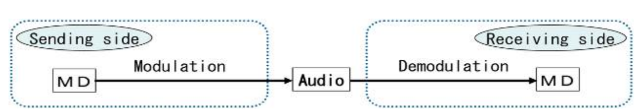


Fig. 1(b) Modulation/Demodulation method

2) Narrow frequency band of the MD signal: The maximum modulation signal frequency for AM broadcasting, which primarily targets sound transmission, is approximately 5 KHz, and is approximately 20 KHz for FM. In comparison with these, the present transmission method is aimed at the transmission of robot MD, namely, joint angle command variables, and thus the maximum frequency is incredibly low at DC to a few hertz. Based on these facts and (1), the carrier frequency region does not initially contain the signal bandwidth and so the transmission of MD as an unmodified sound is not possible. In other words, the baseband transmission method is impossible in principle and modulation/demodulation is indispensable.

3) Simplification of demodulation processing: Typical radio receivers employ an analog circuit for demodulation, and thus the temporal lag between the sound signal at the transmitter and the sound signal at the receiver can be ignored. Meanwhile, due to the recent progress of devices, demodulation processing can sometimes be implemented by digital circuits or software rather than by analog circuits. In these cases, a lag between sound transmission and reception naturally occurs due to the processing time; however, the listener simply listens to the sound and with the exception of time announcements, this basically does not cause any obstacle. Nonetheless, in the case of MD transmission, this delay is fatal. That is, ideally, the MD emitted at the transmitter should be instantaneously transmitted to the robot, which is expected to be replayed and become an actual movement with as little delay as possible. The reason for this is that such a delay causes a significant feeling of discomfort in the people watching the robot content. Moreover, in this research the sound is not only a transmission medium for MD and but also an audio signal for humans to hear. From the above considerations, it is necessary to implement the demodulation processing for this method in analog, or in the case of digital, make the demodulation processing time as short as possible.

4) Number of transmission channels depending on the degrees of freedom of the robot: General communication robots have various degrees of freedom. In the case of humanoid robots, several dozen degrees of freedom are necessary. The intended gesture is correctly realized by synchronizing the MD for each joint of the robot. Accordingly, it is necessary to have the same number of transmission channels and degrees of freedom; furthermore, as the number becomes larger, the demodulation processing becomes more complex. Thus, there is a trade-off between (3) and the robot's degrees of freedom.

As above, because the requirements for the proposed method are extremely different from those for conventional radio broadcasting, its design theory is also different from the conventional one. In other words, it is necessary to optimize (3) under the conditions of (1), (2), and (4). Hereinafter, for simplicity, we shall focus the discussion on only the wired transmission of the audio signal via a sound cable and its digital implementation.

B. Proposed method.

From the above considerations, a frequency-division multiplexed amplitude modulation method on a robot joint basis is proposed in this paper as one of the most simple analog modulation/demodulation methods (Fig. 2). As shown in this figure, the sound data after modulation is as follows.

$$x(t) = \sum_{i=1}^{C} \{k_i(t) + |k_{\min}|\} \cos \omega_i t \quad (1)$$

Here, C is the number of robot joints, $k_i(t)$ is the MD that should be passed to joint i, $k_{\min} < 0$ is the minimum value of $k_i(t)$, ω_i is the angular frequency of the carrier wave for the joint.

An outline of the proposed method is explained as follows. First, the frequency band of the audio signal is determined for each joint at both the transmitter and receiver. The sound of each frequency band is modulated with respect to the MD of the corresponding joint at the transmitter. The signals are superposed and sent. All the signals are superposed and sent. When performing demodulation at the receiver, the sound data $k_i(t) + |k_{\min}|$ is extracted by a band pass filter (BPF) corresponding to ω_i and $k_i(t)$ are restored by performing asynchronous detection.

The structure shown in Fig. 2 resembles the conventional AM radio broadcasting method; however, consideration of $|k_{\min}|$ with respect to condition (2) of Section Ⅳ.A is necessary. Here, consideration is given to the stability of the system and the ease of implementation concerning the BPF for demodulation and the low pass filter (LPF) within the asynchronous detection, and the design of the finite impulse response filter is performed using the window method.

In addition, analog amplitude modulation rather than digital amplitude modulation is employed considering our system environment. As wired transmission is taken up for this method, it is not particularly necessary to consider noise resistance characteristics, and furthermore analog data can better reduce demodulation processing time.

C. Parameter design for demodulation.

The main technical issue in regard to the proposed method

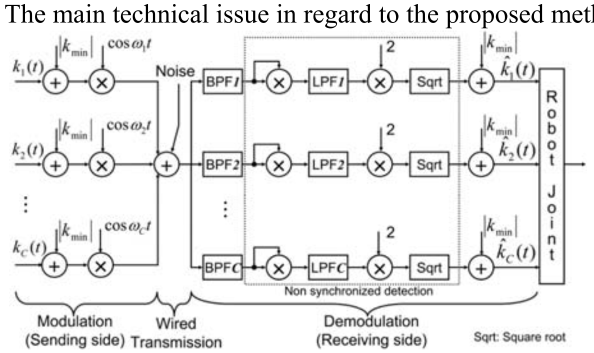


Fig. 2 Motion data transmission flow with AM

is the handling of condition (3). In the demodulation, only the BPF and the LPF during asynchronous detection are variables, and the processing time for both filters must be minimized in order to shorten the overall processing time. The processing time for each filter is roughly proportional to the number of delay devices used in the filter, so the objective is to minimize the number of delay devices in both filters.

Accordingly, a demodulation processing parameter design method that minimizes the demodulation processing time is proposed in regard to the requirements of each system environment. The demodulation parameter design algorithm for the shortest processing time is shown in Fig. 3. Also, the relation of BPF and LPF filter characteristics with the parameters defined in the algorithm are shown in Fig. 4.

The sampling frequency f_s determined depending on the system environment, the maximum frequency of robot motion f_m, the number of transmission channels (= number of robot joints) C, and the allowable interference gain in relation to other joints for the MD of the target joint A are input into the design flow, and other parameters are determined according to the processing flow. Finally, the carrier wave frequency decided between the transmitter and receiver $f_{d1} \sim f_{dC}$, the BPF filter coefficient K_B, and the LPF filter coefficient K_L are output. The calculation formula for the number of delay devices differs according to the allowable interference gain A, and the calculation formula for a gain of 0. 01 is given in Fig. 3. (The calculation formula for the number of delay devices in regard to the allowable interference gain is shown in Table.1 [8]). From Table.1 it can be seen that broadening the transition bandwidth reduces the number of delay devices in the case that the sampling frequency is fixed.

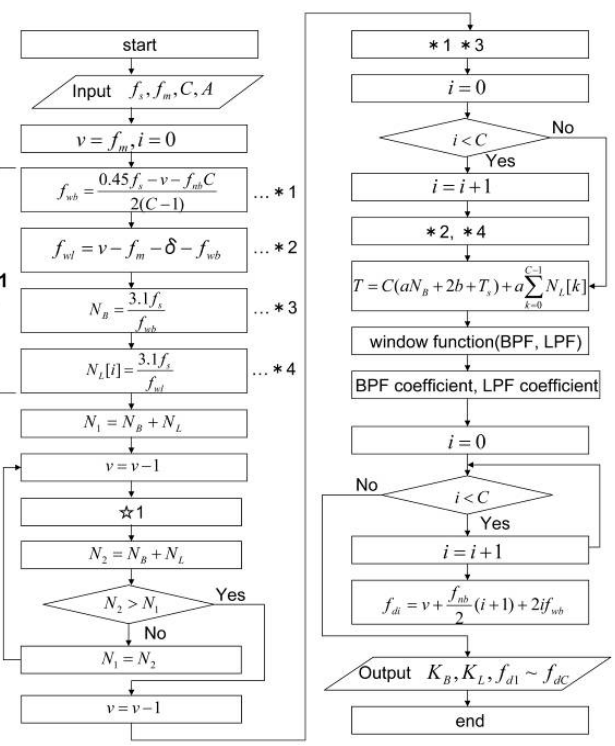
Next, the process of the design flow will be explained. First, the value of the parameter (≥100 f_m) for which the sum of delay devices for BPF and LPF is a minimum is determined for the lowest carrier wave frequency. After determination of v, the transition bandwidth and number of delay devices for the BPF, and the transition bandwidth and the number of delay devices for the LPF for each channel are sought. This is because channels corresponding to higher-frequency carrier waves take a broad transition bandwidth for the LPF and enable a reduction in the number of delay devices. Finally, each filter coefficient is derived by a window formula corresponding to the frequency response and allowable interference gain for both determined filters.

V. TEST SYSTEM

A test system was constructed in order to verify the validity of the proposed method (Figs. 5 and 6). In this study, aiming to apply this method to an automatic presentation system, we implemented the processes for both transmission and reception on a single PC. For this reason, a configuration was adopted where the speaker jack and mike jack were directly connected by an audio cable and the output sound was recaptured within the PC. The IP RobotPHONE was used as the robot, the MD output destination [9].

Fig.5 shows experimental results for the transmission and reception of three-degree-of-freedom data. Here the parameters used in the experiment are as in Table.2.

Fig.5 shows the audio signals modulated using equation (1) at the transmitter. From the same figure, it can be seen that the audio signal is almost exactly transmitted. However, the demodulated signal is approximately 40 ms delayed in comparison with the modulated signal. Upon analyzing the cause of this delay it was determined that 20 ms is required for accumulation of received sound data in the buffer and 20 ms for demodulation processing. Akaida et al. state that the detection limit is about 50 ms in the case that sound is faster than the video in a clip [10]. If this standard also applies to motion, the lag between sound and motion is within the detection limit and can be interpreted as being within the tolerance level.



C : number of channels, A : interference gain,
f_s : sampling frequency, f_m : motion frequency,
f_{wb} : transition band-width(BPF), f_{wl} : transition band-width(LPF),
$f_{d1} \sim f_{dC}$: carrier frequency of each channel,
N_B : number of delay devices(BPF), δ : device error
$N_{L[0]} \sim N_{L[C-1]}$: number of delay device(LPF),
K_B : filter coefficient (BPF), K_L : filter coefficient (LPF),
a, b : variable(T (process time) = aN (number of delay devices) + b),
T : process time(BPF+LPF), T_s : process time(other)

Fig.3 Parameter design chart to minimize the process time

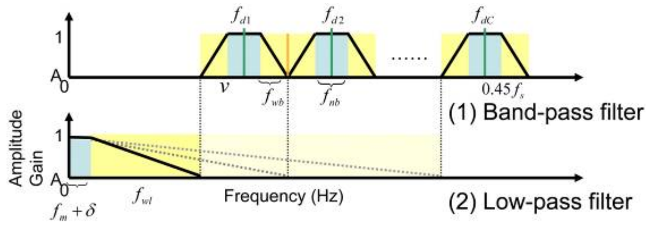


Fig.4 The relationship between filter and parameter

VI. Application Example

A. Necessity of an automatic presentation robot system that links with existing media.

There are many examples in daily life of attempts to automate the tasks of introduction, promotion, and explanation for various products and merchandise. For example, tape recorders and videos play on repeat in supermarkets in order to motivate the customer to make a purchase. However, these are too repetitive and tiresome for the customers, and cannot be said to promote sales greatly. Ideally, it would be most effective for a veteran salesperson to interactively converse with the customer and give a demonstration using the actual product; however, such veteran salespeople are not always available. If some of the functions of a veteran salesperson could be replaced by robots, it can be expected to be useful in sales promotion. Also, another example where a presentation robot could be of service is at poster sessions of academic conferences. The presenters at poster sessions repeat exactly the same presentation many times over as attendees arrive, which is quite burdensome. Accordingly, if an overview of the research could be given by a presentation robot in place of a human, for instance, the human would be able to devote their time and energy to the more important task of fielding questions. Against such a background, a primary objective of this research is development of a robotic system to support human presentations.

Similar concepts have been proposed previously [1-4]. Although the unusual nature of such a robot's is able to attract customers out of curiosity, the information handled by the robot itself is rather poor and the situation is such that robots have not yet reached practical service. Accordingly, in this paper, we consider the possibility of enhancing the effectiveness by compensation for mutual weaknesses through coordination with existing media, not just using a robot alone. As in the previous example of the supermarket, existing media include tape recorders and video tape recorders, as well as personal computers, radio, and television, and even paper posters and the like. While these existing media can handle an enormous quantity of information such as text, sounds, images, and video, their interaction with the real world, such as with customers and real objects, is poor. Accordingly, we considered the coordinated fusion of both groups to be possible by using robots and having them interact with the real world.

B. Proposal of a presentation robot system linked with PowerPoint.

With the aim of making the proposed application widely distributable, attention was given to linkage with Microsoft Office PowerPoint, the de facto standard for presentation software. This application is equipped with a function called "Slide Show", which automates the presentation. However, an API to start up an arbitrary program has not yet been released. Accordingly, attention was given to the automatic

TABLE.1 NUMBER OF DELAY DEVICES AND INTERFERENCE GAIN

Window function	Transition bandwidth(Hz)	Amount of damping (block area) (dB)
Rectangular window	$f_L=\frac{0.9f_s}{N}$	21
Hanning window	$f_L=\frac{3.1f_s}{N}$	44
Hamming window	$f_L=\frac{3.3f_s}{N}$	53
Blackman window	$f_L=\frac{5.5f_s}{N}$	74

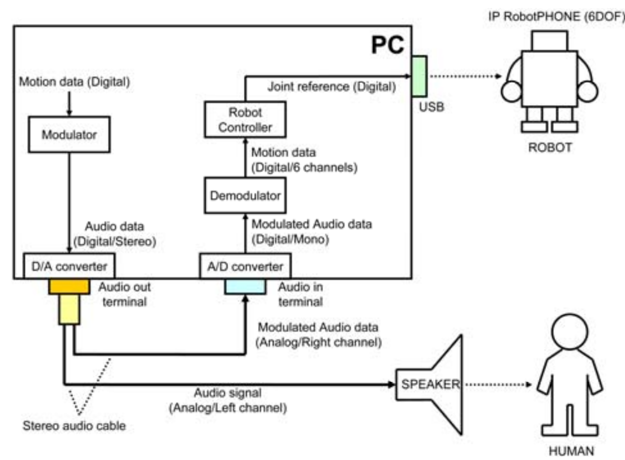


Fig. 5 Block diagram of the experimental system implemented on a single PC


Fig. 6 Overview of the experimental system

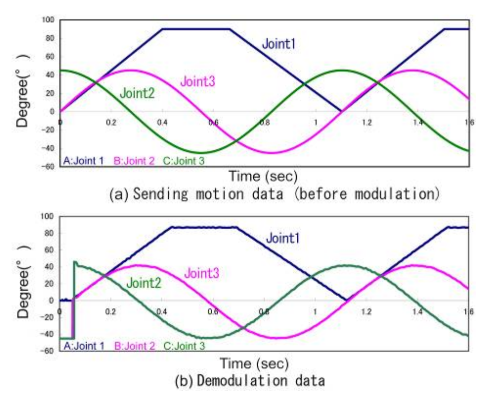


Fig.7 Experimental result

TABLE.2 DESIGNED PARAMETER

Input	Intermediate results	Output
$f_s = 5000$, $f_m = 3$, $C = 3$, $A = 0.01$	$f_{nb} = 18$, $f_{wb} = 301$, $f_{wl} = 673, 1286, 1893$, $N_B = 51$, $N_L[0] = 23, N_L[1] = 11, N_L[2] = 9$	$f_{d1} = 1001$, $f_{d2} = 1612$, $f_{d3} = 2223$

replay function for sound effects in the Slide Show. Specifically we attempt to link the PowerPoint content with a robot by inserting the audio file modulated by the MD into the slide. The sound data here is stereo, and so the sound data that people should hear such as voice or music is allocated to the left channel and the sound data modulating the carrier wave with respect to the MD is allocated to the right channel. We considered a multimedia presentation robot to be feasible, linking the text, image, and sound content within the slide with the robot motion at an arbitrary timing. Fig. 6 shows a schematic diagram of the proposed system. By actually incorporating the MD of equation (1) into the PowerPoint slide, the operation of a robot with the intended timing was empirically confirmed.

C. Robot content example.

An example of content is that a robot uses a pointer to explain the details of a PowerPoint slide. When humans conduct a presentation, they very often perform actions (pointing motions) to control the gaze of the audience, for example, indicating specific locations or items with a laser pointer. By using an audio file modulated with the MD representing the location at which the robot points, it becomes possible for the robot to automatically implement a pointing operation. If such a technique is used, the partial automation of poster sessions such as those mentioned in 6.1 is expected to be facilitated.

Fig.8 shows an example of content in which animation on a PC and a robot are linked. After playing an animation of a character throwing a ball in the slide, MD is transmitted to the robot such that it seems to catch the ball simultaneously

VII. SUMMARY

This paper proposed a frequency-division multiplexed amplitude modulation method and a parameter determination method that gave consideration to minimizing the demodulation processing time as a transmission method for

Fig.8 Collaboration between PowerPoint and a robot

multi-degree-of-freedom motion data by an audio signal. In PowerPoint was proposed and an example of the content introduced.

A precise pointing method at the display or real objects, content creation methods, and expansion of the research to projectors and posters are planned as tasks for the future.

REFERENCES

[1] Mitsubishi Heavy Industries, “Wakamaru presentation system” (in Japanese),http://www.mhi.co.jp/products/detail/wakamaru_presentation.html.

[2] Nishiyama H., et al., “Design of a service robot system based on GRID computing” (in Japanese), 18th Annual Conference of the Japanese Society for Artificial Intelligence, 3G2-04, 2004.

[3] Shiomi M., et al., “Interactive humanoid robots for a science museum”, IEEE Intelligent Systems, vol. 22, no. 2, pp. 25-32, Mar./Apr. 2007.

[4] Byung-Ok Han et al., “Museum tour guide robot with augmented reality”, 16th Virtual Systems and Multimedia, 2010.

[5] Nakayama A., et al., “Rich communication with audio-controlled network robot. Proposal of "Audio-MotionMedia"”, Robot and Human Interactive Communication, 2002.

[6] Gerasimov V., et al., “Things that talk: using sound for device-to-device and device-to-human communication”, IBM Systems Journal 39(3, 4), 530-546, 2000.

[7] Lopes C.V., et al., “Acoustic modems for ubiquitous computing”, IEEE Pervasive Computing 2(3), 62-71, 2003.

[8] Kawamata M., “Foundations of digital signal processing learned in MATLAB: Part 4. Design of FIR files” (in Japanese)

[9] Dairoku Sekiguchi et al., “The Design of Internet-Based RobotPHONE”, NHK Technical Research Institute, ICAT 2004.

[10] Akaida T., et al., “Lip-sync – image and voice timing –” (in Japanese), NHK Technical Research Institute, Mar. 1997.

[11] Owa S., et al., “Robot presenter” (in Japanese), Society of Instrument and Control Engineers, Motion Media Contest, 2009.

[12] Kitagishi Ikuo et al., “Development of Motion Data Description Language for Robots Based on eXtensible Markup Language”, Journal of Robotics and Mechatronics 14(5), 471-478, 2002.

[13] Nishimura Yoshitaka et al., “Development and psychological evaluation of Multimodal Presentation Markup Language for Humanoid Robot”, 5th IEEE-RAS Humanoid Robots, 2005

[14] Hirokazu Konishi. et al., “A study of the character media robot cooperating with TV programs” (in Japanese) , FIT 2002(3), 471-472, 2002.

[15] Owa S. et al., “Research on a multi-degree-of-freedom motion data transmission method using sound” (in Japanese) , ROBOMEC, 2009.