2015 International Conference on Computational Science and Computational Intelligence

One-to-one audio guidance system using human vision, designed for a guide robot

Kenichiro Yasui, Ukai Masashi, Wei Quan, Xiuming Wu, Hideo Furuhashi
Aichi Institute of Technology
1247, Yachigusa, Yakusa-cho, Toyota, JAPAN
v14715vv@aitech.ac.jp

***Abstract*- In this study, we present a guide robot that speaks to humans through a directional loudspeaker. The robot is equipped with an image-sensing device (human vision component) that acquires the age, gender, and direction of the face. The robot rotates in the direction of the human face and speaks; the contents of the speech depend on the information received from the human vision component. The directional loudspeaker is a parametric speaker installed with an ultrasonic speaker array. The system was constructed using the RT middleware. The robot can speak to a single person within a narrow audible area. This paper discusses the characteristics of the developed system.**

Keywords- guide robot; Directional loudspeaker; RT middleware; human vision.

I. INTRODUCATION

In recent years, guide robots that react to human conversations or voice operations have been developed. Guide robots are used in environments such as homes, shops, and museums. In shops and museums, where many people occupy a small area, the audio output should be limited to the person being addressed.

Moreover, interpersonal communications and their contents depend on the age, gender, and other attributes of the speakers.

We have developed a guide robot using a directional loudspeaker that recognizes human faces. The voice of the robot can be heard only by the person being addressed. The contents of the conversation and the speaking manner are altered to suit the age and gender of the person. The characteristics of our proposed system are discussed.

II. EXPERIMENTAL

Figure 1 shows the experimental setup of the system. The system components are two servo motors, a human vision component, and a directional loudspeaker. The motors rotate the vision component and the speaker along the yaw and pitch axes. The human vision component is an imaging sensor (OMRON Corp.) that obtains the age, gender, direction of a face, and other attributes from the facial information. The directional loudspeaker is an ultrasonic parametric speaker (Tristate Ltd.) installed with 50 ultrasonic transceiver elements (5 in the vertical direction, 10 in the horizontal direction).

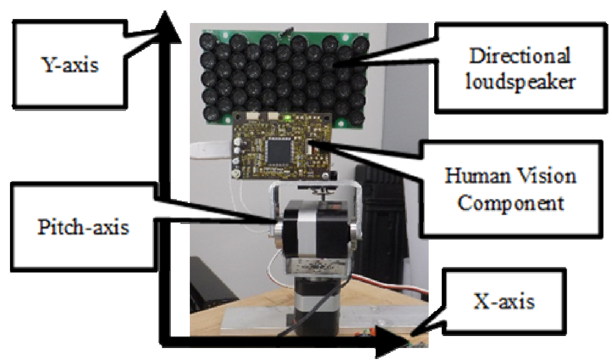
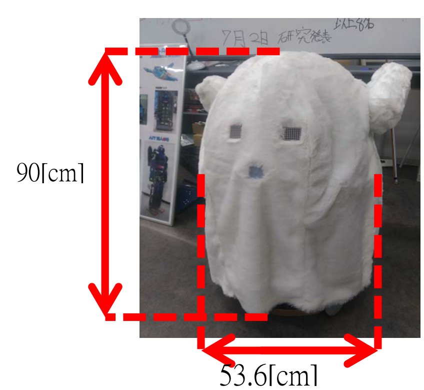


Fig. 1 Experimental setup

Figure 2 presents the communication robot containing the developed system. The robot is 90 cm high, 53.6 cm wide, and weighs 18 kg. It moves in all directions on omni wheels. Its maximum speed is 1.4 m/s. The omni wheels are used to control the equipment for yaw axes.



978-1-4673-9795-7/15 $31.00 © 2015 IEEE
DOI 10.1109/CSCI.2015.40

Fig. 2 Communication robot

III. SYSTEM

An overview of the system design is shown in Fig. 3. The program is constructed using RT middleware (Open RTM-aist). Each function is constructed as one component. First, the human vision component outputs the age, gender, and direction of the face extracted from camera images. Second, the motor directs the system toward the face. Third, the contents of the speech are determined from the age and gender of the person. Finally, the contents are output by the directional loudspeaker.

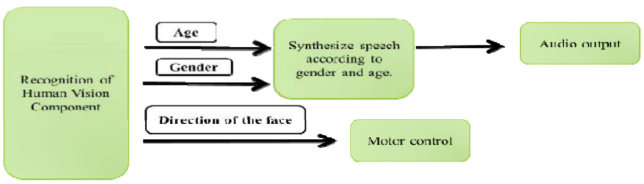


Fig. 3 Overview of the system design

IV. EXPERIMENTAL

The directivity of the directional speaker was measured at 1 m from the speaker. The test signal was a 1 kHz audio signal. The results are plotted in Fig. 4.

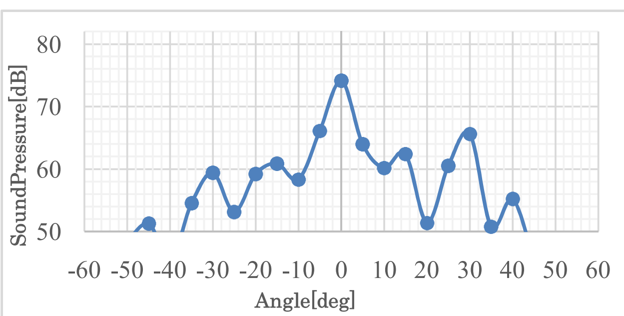


Fig. 4 Directivity of the speaker with a 1 kHz signal

The maximum sound pressure is 75 dB. The half-width-at-half maximum (−6 dB) of the pressure is 4°. The distinctive distance of the speaker is 15 cm.

Next, the system was installed in the robot, and a person was positioned at 1 m and 30° orientation from the front of the robot. Once the system was started, the human vision component found the face and the motors guided the system toward the face. The guide contents were determined on the basis of the information of the human vision component and were output by the speaker. Figure 5 plots the directivity of the sound measured at 1 m from the robot.

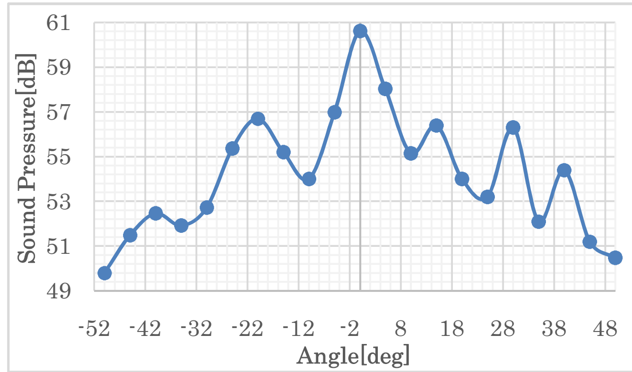


Fig. 5 Directivity of the speaker outputting the guide contents.

The sound pressure is maximized at the front of the system and the directivity is similar with that of the 1 kHz signal shown in Fig. 4. Therefore, the system communicates with a specific person and can alter its guide contents.

V. CONCLUSIONS

We have developed a guide robot using a directional loudspeaker that responds to human face information. The image-sensing device (human vision component) acquires the age, gender, and angle of the face. The robot can communicate with a specific person within a narrow audible range.

In the next step, we will attempt to control multiple directional speakers that simultaneously address a larger audience.

ACKNOWLEDGMENT

This work was supported by JSPS Grant Number 15K05879.

REFERENCES

[1] K Yasui "Audio guidance system in specific person for a guide robot", Tokai section conference on Electrical, Electronics, Infomation, and Related Engineering Japan, (2015).

[2] K Nakadai "Towards New Human-Humanoid Communication: Listening During Speaking by Using Ultrasonic Directional Speaker", Honda Research Institute Japan, (April 2005)