International Journal of Automation and Computing 16(1), February 2019, 93-101
DOI: 10.1007/s11633-016-1022-2

Virtual Plate Based Controlling Strategy of Toy Play for Robot′s Communication Development in JA Space

Wei Wang Xiao-Dan Huang
School of Information and Electrical Engineering, Hebei University of Engineering, Handan 056038, China

Abstract: Toy play is a basic skill for a humanoid robot after it has joint attention (JA) ability. Because such skill is helpful for human-robot interaction and cooperation, we must realize this skill to enhance the robots communication ability with person. In this paper, we researched a toy play controlling strategy in JA space based on a virtual plate with a serial robot arm, which has five degrees of freedom (5-DoF). For this purpose, a reachable space of joint attention was constructed firstly. And then the toy play controlling strategy was proposed in details. Here we used a virtual plate to enhance the toy play effect. In order to realize this skill better, toy play energy and some restraining relations were analyzed. By contrasting the audio waveform in the experiments, good performance effect of toy play was demonstrated.

Keywords: Human robot cooperation, joint attention (JA) space, reachable space, toy play ability, a virtual plate.

1 Introduction

Various theoretical accounts propose that an important developmental relation exists between joint attention, play, imitation abilities, and communication ability[1−3]. Take young children for instance, joint attention and immediate imitation are strongly associated with language ability, whereas toy play and deferred imitation are the best predictors of rate of communication development. When a child can focus his attention on his partners, he learns and imitates actions from his mate. Toy play, as a positive feedback, can consolidate two abilities above. Moreover, it provides a context for the development of communication ability.

On the other hand, with the theoretical and technological development of robots, researchers pay more attention on human robot interaction and cooperation. Most robots have basic interaction ability. Some of them can speak, watch or act, and others have affection[4, 5]. But in order to make them be more popular and widespread in our life, intelligence and sociality should be focused. Intelligence is from mind, and sociality is from communication ability. Therefore, toy play also takes an important role for robots researches and development.

To endow a robot with intelligence and sociality, and carry on whole researches about joint attention, play, imitation abilities, and communication ability for the robot, we developed a toy play unit in our joint attention system and studied a related controlling strategy. Based on a body coordinate system of robot, we construct a reachable space. In this space, the robot plays instrument on virtual instrument plate with a controlling strategy. To ensure the effect of the method, the striking energy and restraining condition between the position of two holders and the distance of the two plates are analyzed.

The overall goal of our work is to explore and implement an effective controlling strategy for a toy play unit in the joint attention system. This paper is structured as follows. Firstly, in Section 2, related works on joint attention for robots are stated. Secondly, in Section 3, we introduce our toy play unit in the joint attention system. Moreover, toy play controlling strategy is proposed based on a reachable space of joint attention. Thirdly, some experiments are done and results are analyzed in Section 4. Finally, conclusions are given in Section 5.

2 Related works

Joint attention (JA) is an early social cognitive ability, which refers to one individual can follow another individual′s attention to make both of them pay attention to the same object at the same time. It is also called "shared-attention" in Taiwan[6]. This process has many expressing styles, such as pointing, acting or gazing. Take gazing for example shown in Fig. 1.

Fig. 1 shows the whole process of joint attention in gaze style. Firstly, a person stares at something exhibiting attention. Then, a robot obtains this information and also shows attention to that thing. So sharing attention is the foundation of cooperation between human and robot to complete the same task. For the problems about joint attention in

Research Article
Manuscript received September 7, 2015; accepted February 22, 2016; published online February 21, 2017
Recommended by Associate Editor Li-Hui Wang

human robot cooperation, the United States, Germany, Britain, Japan, conducted more researches and had some achievements[7−11]. Studies on JA are divided into three categories. They are responding to joint attention (RJA), initiating joint attention (IJA), and ensuring joint attention (EJA)[7]. In the RJA study, Carlson[12, 13] proposed a learning model for developing joint attention ability. Breazeal[14] developed a JA functional module. In IJA aspect, Imai et al.[15] studied joint attention system with vision . And reference [7] also conducted a study on the combination of EJA, RJA and IJA. In addition, a method that a robot learns JA ability was introduced in [16]. And reference [17] solved a constrained problem, i.e., when to change the direction of the line of sight which must be specified in the robot interactive process by bringing in attention selecting mechanism. Hashimoto et al.[18] in Japanese Shinshu University realized JA ability on a robot head platform. And Chu et al.[19] proposed a visual attention model for robot object tracking. Table 1 compares experiment platforms, research directions and characteristics of JA, respectively.

Many researchers studied the toy play controlling strategy on different platforms and from different points of view. Breazeal[20] presents results illustrating how this control architecture, embodied within an expressive robot and situated in a social environment, enables the robot to socially influence its human care giver into satisfying its goals. With their experiment platform, Kismet has a variety of affective responses when interacting with different persons or colorful toys. Leon[21] selected a toy, yo-yo, for their two control strategies. One based on predefined hand motion pattern and the other generating the hand motion on-line. Both allow playing the yo-yo at a selected top height. Petric et al.[22] presents a novel method to obtain the basic frequency of an unknown periodic signal with an arbitrary waveform, which can work online with no additional signal processing or logical operations. The proposed method can be used for the control of rhythmic robotic tasks, where only the extraction of the basic frequency is crucial. Cook et al.[23] reviewed assistive robots for playing, learning and cognitive development.

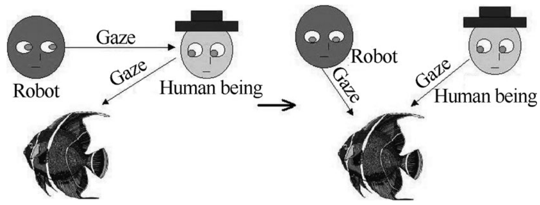


Fig. 1 Sketch of joint attention in gaze style

Table 1 Platforms and focuses about joint attention

Organization	Platform	Researches
1) Department of Adaptive Machine Systems, Osaka University, Japan, HANDAI Frontier Research Center, Graduate School of Engineering 2) National Institute of Information and Communications Technology, Kyoto, Japan 3) Department of Computer Science, Yale University, USA	1) Robot head (3DOF, 2 cameras) Robot motion mechanism (2 cameras, 3DOF), Simulation Platform, Robot agent 2) Infanoid2 (2 cameras, Neck-3DOF,Eye-3DOF) 3) Nico (Upper torso robot, Head-7DOF, Arm-6DOF, Eye-2 CCD cameras)	JA learning mechanisms, models
Saitama University, Japan	Robot head (3D facial mask, LED ProjectorPan unit1 camera)	Control for a specific user′s attention
1) Chiba Institute, Japan, Tokyo University of Electro-Communications 2) Autonomous Intelligent Systems Group, University of Bonn, Germany 3) Department of Mechanical Engineering, National Taiwan University	1) Robovie-R (Bumblebee2 camera) 2) Domestic service robot (Laser rangefinders, Time-of-Flight camera) 3) Pan unit	JA image processing related research
1) Communications Research Laboratories in Kyoto, Department of Linguistics, University of Lund, Sweden 2) Bielefeld University, Institute of Technology, Applied Computer Science	1) Infanoid1 (Upper torso robot, 24DOF, 480mm high4CCD cameras) 2) BARTHOC (Face-10DOF2 Color cameras, Arm)	JA as one interactive feature, interactive multi-channel integrated research
ECE Department, University of Waterloo, Canada, College of Engineering, Memorial University	T265 CRS ARM (Point Grey Research Flea2 camera, Pan unit, Laser pointer)	Bionic artificial visual attention model
Georgia Institute of Technology	Simon(Upper torso robot, Arm-7DOF, Hand-4DOF, Eye-2DOF, eyelids, color ears)	RJA+IJA+EJA

In the related literature, we researched on JA theories and technologies deeply. Based on a 5-DOF serial robot arm with a fixed monocular vision system, joint attention space was analyzed. And then RJA is realized with coordinate transformation technology. It is the foundation for further researches about toy play, imitation abilities, and communication ability. In this paper, the intention of this research is to develop a toy play unit in joint attention systems and study the related controlling strategy.

3 Toy play controlling strategy

3.1 Toy play unit in joint attention system (JAS)

The toy play unit in a joint attention system is a double-loop system with a joint servo control loop and a vision control loop based on a fixed monocular vision system. It consists of a camera, a PC, a control board, wireless modules and a serial robot arm with five degrees of freedom, as shown in Fig. 2.

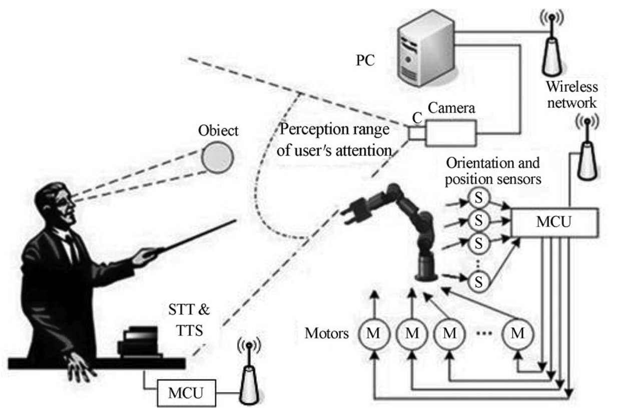


Fig. 2 Hardware platform of toy play unit in joint attention systems

3.2 Reachable space of joint attention

A joint attention space includes an observable space and a reachable space. In the observable space, user attention could be captured accurately, and in the reachable space, human robot cooperation could be carried on. Because the toy play is mainly realized in the reachable space, we state its constructing process in detail as follows.

3.2.1 Body coordinate system of robot arm

The rear coordinate system of the robot arm is established[24]. There are six coordinate systems in total. They are $O_0X_0Y_0Z_0$, $O_1X_1Y_1Z_1$, $O_2X_2Y_2Z_2$, $O_3X_3Y_3Z_3$, $O_4X_4Y_4Z_4$, $O_5X_5Y_5Z_5$, as shown in Fig. 3.

Based on the established Denavit-Hartenberg (D-H) coordinate system of the robot arm, D-H parameters and the joint variables of each link are shown in Table 2.

Table 2 D-H parameter of arm

Joint i	α_i(degree)	a_i(mm)	d_i(mm)	θ_i(degree)
1	−90	0	38	$\theta_1\in$[−90, 90]
2	0	154	0	$\theta_2\in$[−180, 0]
3	0	122	0	$\theta_3\in$[−30, 150]
4	−90	0	0	$\theta_4\in$[−180, 0]
5	0	0	181	$\theta_5\in$[−90, 90]

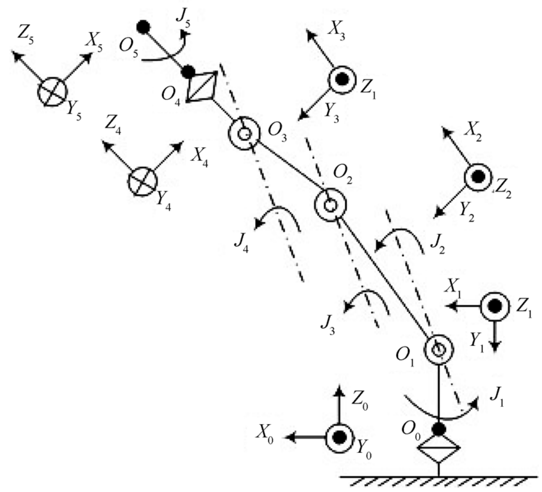


Fig. 3 D-H coordinate of the robot arm

According to the parameters above and D-H formulas:

$$ {}^0_1T = \begin{bmatrix} c\theta_1 & 0 & -s\theta_1 & 0 \\ s\theta_1 & 0 & c\theta_1 & 0 \\ 0 & -1 & 0 & d_1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{1} $$

$$ {}^1_2T = \begin{bmatrix} c\theta_2 & -s\theta_2 & 0 & a_2\times c\theta_2 \\ s\theta_2 & c\theta_2 & 0 & a_2\times s\theta_2 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2} $$

$$ {}^2_3T = \begin{bmatrix} c\theta_3 & -s\theta_3 & 0 & a_3\times c\theta_3 \\ s\theta_3 & c\theta_3 & 0 & a_3\times s\theta_3 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{3} $$

$$ {}^3_4T = \begin{bmatrix} c\theta_4 & 0 & -s\theta_4 & 0 \\ s\theta_4 & 0 & c\theta_4 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{4} $$

$$ {}^4_5T = \begin{bmatrix} c\theta_5 & -s\theta_5 & 0 & 0 \\ s\theta_5 & c\theta_5 & 0 & 0 \\ 0 & 0 & 1 & d_5 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{5} $$

where s denotes $\sin(\cdot)$ and c denotes $\cos(\cdot)$.

Multiply each link matrix to obtain the posture and orientation of the end effector of the robot arm in the world coordinate system.

$$ {}_5^0T = {}_1^0T \times {}_2^1T \times {}_3^2T \times {}_4^3T \times {}_5^4T. \qquad (6) $$

Thus, by solving inverse kinematics, joint variables of the robot arm can be calculated with the posture and orientation of the end effector.

3.2.2 Construction of the reachable space

Reachable space of joint attention refers to a maximum range of activities of a reference point in the wrist mechanical interface coordinate system (the end effector coordinate system). Without considering the restrictions for the joint angle, for rotary joints, a fixed reference point P_n is assumed on the robot arm. It rotates around Z_n axis with the end coordinate system $O_nX_nY_nZ_n$ together. And it forms a joint attention reachable sub-space $W_{n-1}(P_n)$ in the coordinate system $O_{n-1}X_{n-1}Y_{n-1}Z_{n-1}$, which is a circle. Then, the movement of linkage n-1 drives $W_{n-1}(P_n)$ rotating around Z_{n-1} axis form another joint attention reachable sub-space $W_{n-2}(P_n)$ in the coordinate system $O_{n-2}X_{n-2}Y_{n-2}Z_{n-2}$, which is a toroid. After that, the movement of linkage $n-2$ drives $W_{n-2}(P_n)$ rotating around Z_{n-2} axis to form another joint attention reachable sub-space $W_{n-3}(P_n)$ in the coordinate system $O_{n-3}X_{n-3}Y_{n-3}Z_{n-3}$, which is a spinning body. Rotate it around the former axis continually, the joint attention reachable sub-space of the fixed reference point P_n is also a spinning body. There is a transformation between neighboring sub-spaces:

$$ W_{n-j-1}(P_n) = Rot(z_{n-j}, \theta_{n-j}) \times W_{n-j}(P_n) \qquad (7) $$

where $j = 0, 1, \cdots, n-2$.

According to the structure and parameters of the robot arm, $n = 5$. As the statement above, we can get the reachable space of joint attention in human robot cooperating process shown in Fig. 4.

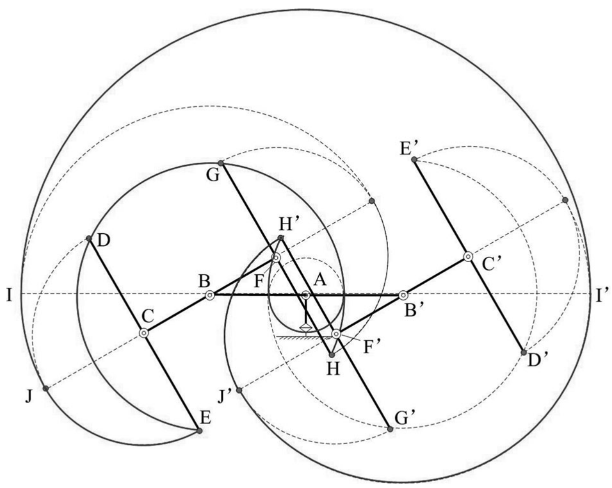


Fig. 4 Construction of the reachable space of the robot arm

3.3 Instrument playing based on virtual instrument plate

The robot arm plays an instrument in the reachable space constructed above. Such system is shown as Fig. 5. Because melodious voice is produced by rhythmic vibrations, the disturbed changes in voice amplitude make it awful and teasing. In the process of instrument performance, the time of hammer striking on an instrument plate plays an important role. The longer it is, the worse it vibrates. However, for the robot arm has inertia, shortening the striking time by controlling motors is not easy for the motors must have higher performance. So a method based on a virtual instrument plate is proposed in this paper. Opening and closing the end effectors control the hammer striking to realize a better performance.

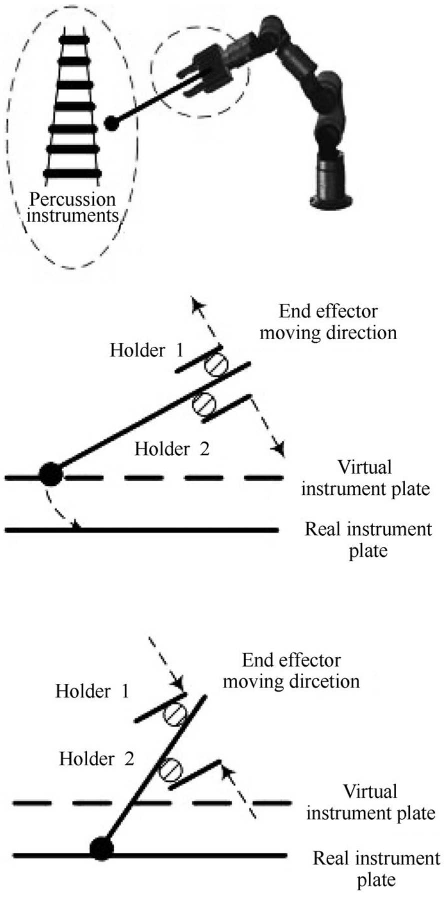


Fig. 5 Instrument performing process

In the reachable space of the robot arm mentioned above, the front end of the hammer can move to a desired point. Suppose that there is a virtual instrument plate ψ above the real instrument plate ξ as shown in Fig. 5. Before striking, the front end of the hammer moves to the virtual instrument plate ψ along with the robot arm. Then, the end effector opens. And the Holders 1 and 2 fixed on the end effector moves relatively. Because of the gravity, the front end of hammer falls down freely around the Holder 2 to the real instrument plate ξ . Finally, the end effector closes to make the front end of hammer revert to the virtual instrument plate. This is the whole striking process.

Given the method proposed in this paper, the performing effect is related with two key points. One is striking energy, and the other is the restraining relationship between the position of the two holders and the distance of the two plates. The statement below will analyze them separately.

3.3.1 Analysis of striking energy

Suppose the distance between Holder 2 and the front end of hammer is l_1, the distance between Holder 2 and the terminal of hammer is l_2, and the distance between Holders 1 and Holder 2 is l_3. The height between the real instrument plate and the virtual one is h. At the beginning, the robot arm holds the hammer and makes its front end to the virtual instrument plate. The angle of the hammer and plate is α. And when the front end of hammer falls to the real instrument plate, the angle changes to β, as shown in Fig. 7.

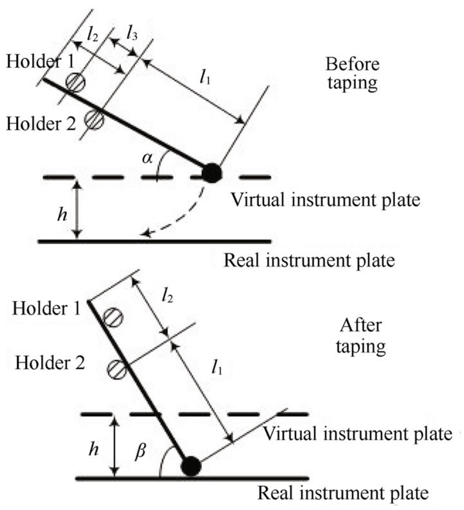


Fig. 6 Instrument performing process

The relation between the heights of real and virtual instrument plates h and the distance of the end effector d is analyzed as follows. As shown in Fig. 6, Holders 1 and 2 move relatively. So the front end of the hammer falls down around the center of them. The trajectory is a circle with the center O. From Before-taping position to After-taping position, the moving details of the terminal of the hammer are shown in Fig. 7.

$$\beta-\alpha=\arctan\left(\frac{2l}{l_3}\right). \tag{8}$$

For $l \approx d$,

$$\beta-\alpha\approx\arctan\left(\frac{2d}{l_3}\right). \tag{9}$$

Simultaneously,

$$h_1=d\times\frac{l_2-\frac{l_3}{2}}{\frac{l_3}{2}}=d\times\frac{2l_2-l_3}{l_3}. \tag{10}$$

We can deduce that

$$h_2=h_1\times\frac{l_1+\frac{l_3}{2}}{l_2-\frac{l_3}{2}}=h_1\times\frac{2l_1+l_3}{2l_2-l_3}. \tag{11}$$

Because of $\gamma=\pi-\frac{(\beta-\alpha)}{2}$,

$$l_4=\frac{h_2}{\sin\gamma}=\frac{h_1}{\frac{\cos(\beta-\alpha)}{2}}\times\frac{2l_1+l_3}{2l_2-l_3} \tag{12}$$

$$\eta=\pi-\beta-\gamma=\pi-\beta-\frac{\pi-(\beta-\alpha)}{2}=\frac{\pi-\beta-\alpha}{2}. \tag{13}$$

So,

$$h=l_4\times\sin\eta=\frac{h_2}{\sin\gamma}\times\sin\eta=\frac{h_1}{\frac{\cos(\beta-\alpha)}{2}}\times\frac{2l_1+l_3}{2l_2-l_3}\times\sin\frac{\pi-\beta-\alpha}{2}=\frac{d\times(2l_1+l_3)}{\frac{l_3\times\cos(\beta-\alpha)}{2}}\times\frac{\cos(\beta+\alpha)}{2}. \tag{14}$$

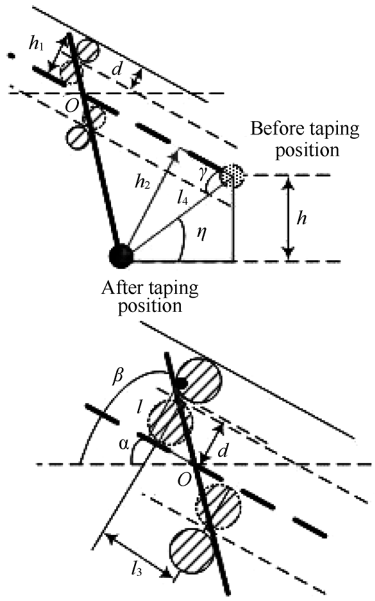


Fig. 7 Moving details of the terminal of the hammer

From (14), we can deduce that h is not related to l_2, so the length between Holder 2 and the terminal of hammer does not affect the height of the real and virtual instrument plates. While the mass of hammer is m, and is almost concentrated at the front end, so the striking energy is

$$E=mgh=mg\times\frac{d\times(2l_1+l_3)}{\frac{l_3\times\cos(\beta-\alpha)}{2}}\times\frac{\cos(\beta+\alpha)}{2}. \tag{15}$$

Moreover, rate of change of energy is considered. Suppose that the distance between the end effectors is Δd, the initial values of l_1 and l_3 are $l_1^{(0)}$ and $l_3^{(0)}$ before knocking, the changing distances of l_1 and l_3 are Δl_1 and Δl_3 after knocking, respectively. Rewriting the equation above into an incremental form, the knocking energy change caused by the changing distance of the end effectors is shown as follows.

$$\frac{\Delta E}{\Delta d}=\frac{mg\Delta h}{\Delta d}=$$
$$mg\times\frac{2\times(l_1^{(0)}-\Delta l_1)+(l_3^{(0)}+\Delta l_3)}{(l_3^{(0)}+\Delta l_3)\times\frac{\cos(\beta-\alpha)}{2}}\times\frac{\cos(\beta+\alpha)}{2}=$$
$$mg\times\frac{2\times(l_1^{(0)}-\Delta l)+(l_3^{(0)}+2\Delta l)}{(l_3^{(0)}+2\Delta l)\times\frac{\cos(\beta-\alpha)}{2}}\times\frac{\cos(\beta+\alpha)}{2}. \tag{16}$$

where $\Delta l=\frac{\Delta d}{\sin(\beta-\alpha)}-\frac{l_3^{(0)}}{2}$, so

$$\frac{\Delta E}{\Delta d}=mg\times\frac{2l_1^{(0)}+l_3^{(0)}}{\Delta d}\times\sin\frac{\beta-\alpha}{2}\times\cos\frac{\beta+\alpha}{2}. \tag{17}$$

It reveals that the energy changing is gradually reduced when the changing distance of the end effectors is increasing.

3.3.2 Analysis of the restraining between the position of the two holders and the distance of the two plates

When the terminal of the hammer is in the Before-taping position, the height between it and the virtual instrument plate is $l_1\times\sin\alpha$. So the height of the virtual and real instrument plate is $\Delta=(l_1+\frac{l_3}{2})\times(1-\sin\alpha)$. In this situation, the falling down distance of the front end of the hammer h must satisfy the inequality $h>\Delta$.

Separately, because motors drive the end effector of the robot arm opening and closing, their rotating angle is limited in 0−180°. For the end effector, there is a maximum changing distance $d_{\max}$. According to (14), the restraining relationship between the distance of the two holders l_3 and the distance of the two plates Δ is as follows.

$$\begin{cases}h>(l_1+\frac{l_3}{2})\times(1-\sin\alpha)\\ \text{s.t. } d\le d_{\max}\\ h\times l_3\times\cos\frac{\beta-\alpha}{2}-2dl_1\times\cos\frac{\beta+\alpha}{2}=dl_3\times\cos\frac{\beta+\alpha}{2}.\end{cases} \tag{18}$$

Suppose that $l_1=\tau\times l_3$ ($\tau\in\mathbf{R},\tau>1$), the restraining above is simplified as follows:

$$\begin{cases}h>\frac{2\tau+1}{2}\times(1-\sin\alpha)\times l_3\\ \text{s.t. } d\le d_{\max}\\ h\times\cos\frac{\beta-\alpha}{2}-2d\tau\times\cos\frac{\beta+\alpha}{2}=d\times\cos\frac{\beta+\alpha}{2}.\end{cases} \tag{19}$$

When parameters $\alpha\in[\frac{\pi}{6},\frac{\pi}{4}]$, $\beta\in[\frac{\pi}{4},\frac{\pi}{2})$, the restraining relationship is displayed in Fig. 8.

In Fig. 8, x-coordinate is τ, y-coordinate is l_3, and z-coordinate is h. When $\tau\in[3,10]$ and $l_3\in[20\,\text{mm}, 34\,\text{mm}]$, the restraining trend is the same as above. But the restraining curve moves along with the changing parameters. α and β get bigger, the restraining curve moves down.

4 Experiments

The whole system consist of a robot arm, a controlling unit, a driving unit, and an instrument as shown in Fig. 9. On the ordinary PC platform (Intel CPU E5700 dual-core 3.0 GHz, 1 G RAM, Windows XP operating system), based on the developing environment Matlab 7.0 and VC++ 6.0, the controlling order of instrument performing is calculated by PC, and drives the robot arm with controlling and driving units.

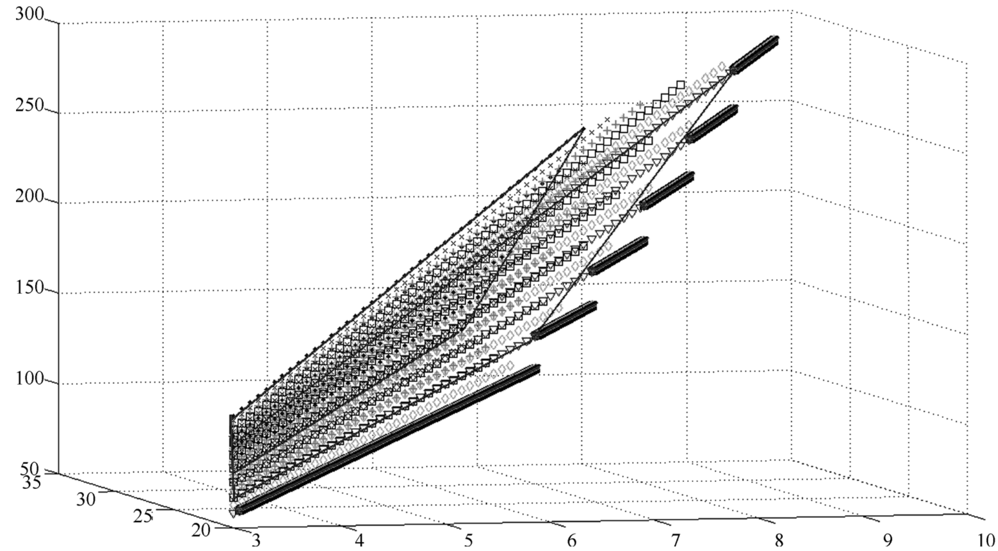


Fig. 8 Restraining between the position of the two holders and the distance of the two plates

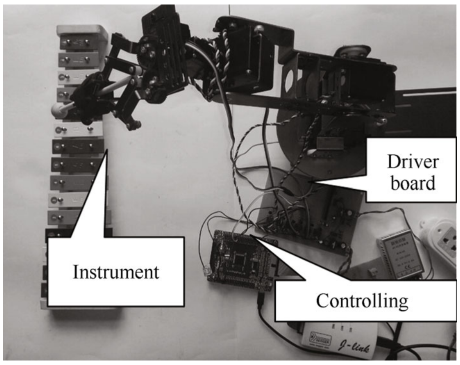


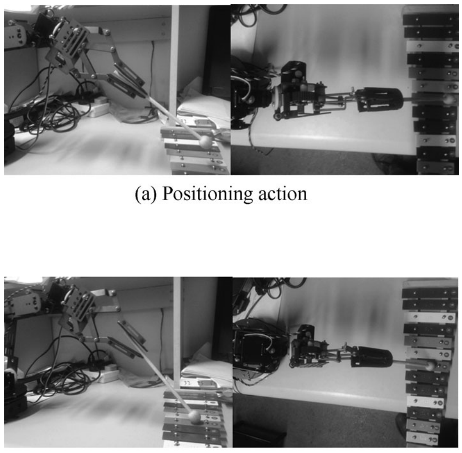
Fig. 9 Toy play unit

4.1 Performing process

Based on the hardware platform and the virtual instrument plate method, three contrasting experiments were conducted. They are single note playing, 2/4 quadruple time music "Doll and Dancing Bear" performing, and 4/4 quadruple time music "twinkle twinkle little star" performing. In each experiment, three striking methods are adopted. They are: 1) the method based on the virtual instrument plate mentioned above, 2) the method that the robot arm drives the hammer directly, and 3) the method that a person drives the hammer. Every striking process is recorded by video and audio. The striking process based upon the method 1 is shown in Fig. 10. Firstly, in the reachable space mentioned above, the robot arm drives the front end of the hammer to the virtual instrument plate. Secondly, the end effectors are made loose to make the front end of the hammer fall down around Holder 2 to the real instrument plate because of the gravity. Finally, the end effectors are tightened to complete the whole striking process, so the front end of the hammer comes back to the virtual instrument plate.

4.2 Performing effects

In order to manifest the effects of the method proposed in this paper, the instrument is struck based on three methods above. Performing effect is contrasted by the waveform of the audio. Firstly, the waveform of single note is shown in Fig. 11. Take Note 1 for instance, there is a gradient changing waveform for the Striking methods 1 and 3. At the moment of striking, the loudness is higher, the energy is stronger, and the amplitude of the waveform is bigger. As time goes by, the loudness and amplitude are decreased to zero. The whole changing process is smooth, so the voice is melodious. For the Striking method 2, the loudness energy and amplitude are decreased as time goes. But the whole changing process is not smooth and there is an abrupt changing area. This is the result that the voice is not melodious. And for other notes, there is the same result too.



(a) Positioning action

(b) Striking action

Fig. 10 Whole striking process based on the method 1

Secondly, the waveform of "doll and dancing bear" and "twinkle twinkle little star" performing is recorded while the music is played based on three methods above. The waveform is shown in Figs. 12 and 13.

The analysis of the music performing process is similar to the statement above. There is a gradient changing waveform for the Striking methods 1 and 3, so the voice is melodious. However, for the Striking method 2, the whole changing process is not smooth and the voice is not melodious.

4.3 Conclusions and future work

1) Toy play is not only a positive feedback, which can consolidate the joint attention and imitation abilities, but also provides a context for the development of later mind/communication ability. So it is a basic skill for a humanoid robot after it has got joint attention ability.

2) In the reachable space of joint attention, an effective controlling strategy of instrument performing based on virtual instrument plate was researched. And striking energy and some constraints were analyzed subsequently.

3) Based on the Robot Toolbox in Matlab, reachable space of joint attention of the robot arm was built. And audio processing software was used to manifest good voice effect by using the method proposed in this paper.

4) Future studies includes: More types of toy play unit in JA space can be researched, such as plucked-string instrument, string instrument and so on, to carry on deep researches about robots joint attention, play, imitation abilities, and later mind/communication ability.

Acknowledgments

This work was supported by Hebei Province Natural Science Foundation for Youths (No. F2015402108), the Foundation for Young Scholars of Hebei Educational Committee (No. QN20131152), Handan Municipal Science and Technology Projects (No. 1421103054).

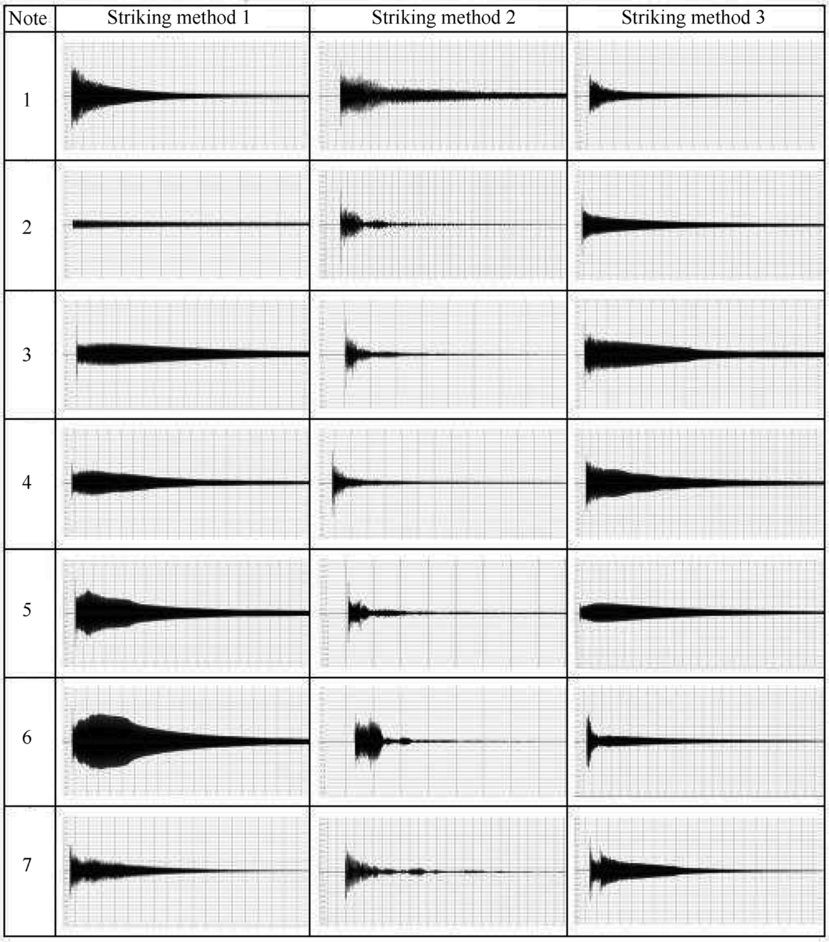


Fig. 11 Waveform for single note based on different striking methods

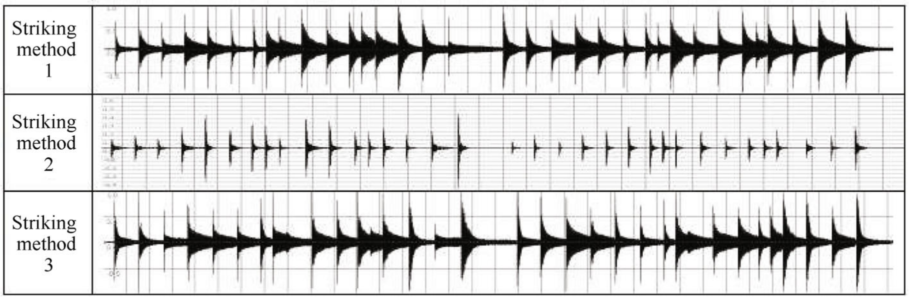


Fig. 12 Waveform for 2/4 quadruple time music based on different striking methods

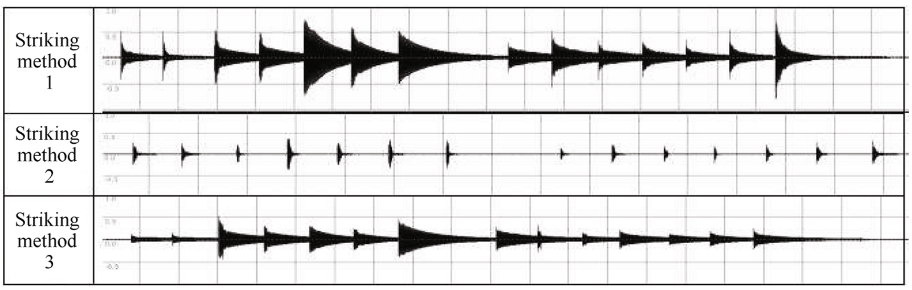


Fig. 13 Waveform for 4/4 quadruple time music based on different striking methods

References

[1] T. Charman, S. Baron-Cohen, J. Swettenham, G. Baird, A. Cox, A. Drew. Testing joint attention, imitation, and play as infancy precursors to language and theory of mind. *Cognitive Development*, vol. 15, no. 4, pp. 481–498, 2000.

[2] S. Dunphy-Lelii, J. LaBounty, J. D. Lane, H. M. Wellman. The social context of infant intention understanding. *Journal of Cognition and Development*, vol. 15, no. 1, pp. 60–77, 2014.

[3] T. Grossmann, S. Lloyd-Fox, M. H. Johnson. Brain responses reveal young infants′ sensitivity to when a social partner follows their gaze. *Developmental Cognitive Neuroscience*, vol. 6, pp. 155–161, 2013.

[4] Z. Y. Xia, L. Li, J. Xiong, Y. Qiang, K. Chen. Design aspects and development of humanoid robot THBIP-2. *Robotica*, vol. 26, no. 1, pp. 109–116, 2008.

[5] L. G. Zhang, Q. Huang, J. Yang, Y. Shi, Z. J. Wang, A. R. Jafri. Design of humanoid complicated dynamic motion with similarity considered. *Acta Automatica Sinica*, vol. 33, no. 5, pp. 522–528, 2007. (in Chinese)

[6] H. Sumioka, Y. Yoshikawa, M. Asada. Reproducing interaction contingency toward open-ended development of social actions: case study on joint attention. *IEEE Transactions on Autonomous Mental Development*, vol. 2, no. 1, pp. 40–50, 2010.

[7] C. M. Huang, A. L. Thomaz. Joint attention in human-robot interaction. In *Proceedings of the AAAI Fall Symposium*, AAAI, Menlo Park, USA, pp. 32–37, 2010.

[8] S. M. Anzalone, S. Boucenna, S. Ivaldi, M. Chetouani. Evaluating the engagement with social robots. *International Journal of Social Robotics*, vol. 7, no. 4, pp. 465–478, 2015.

[9] J. F. Ferreira, J. Dias. Attentional mechanisms for socially interactive robots-A survey. *IEEE Transactions on Autonomous Mental Development*, vol. 6, no. 2, pp. 110–125, 2014.

[10] G. Skantze, A. Hjalmarsson, C. Oertel. Turn-taking, feedback and joint attention in situated human-robot interaction. *Speech Communication*, vol. 65, pp. 50–66, 2014.

[11] T. Ichijo, N. Munekata, K. Hiraki, T. Ono. Entrainment effect caused by joint attention of two robots. In *Proceedings of the 9th Annual ACM/IEEE International Conference on Human-Robot Interaction*, ACM, New York, USA, pp. 178–179, 2014.

[12] E. Carlson, J. Triesch. A computational model of the emergence of gaze following. *Connectionist Models of Cognition and Perception II*, H. Bowman, C. Labiouse, Eds., Singapore: World Scientific, pp. 105–114, 2003.

[13] Y. Nagai, K. Hosoda, A. Morita, M. Asada. A constructive model for the development of joint attention. *Connection Science*, vol. 15, no. 4, pp. 211–229, 2003.

[14] C. Breazeal, D. Buchsbaum, J. Grey, D. Gatenby, B. Blumberg. Learning from and about others: Towards using imitation to bootstrap the social understanding of others by robots. *Artificial Life*, vol. 11, no. 1–2, pp. 31–62, 2005.

[15] M. Imai, T. Ono, H. Ishiguro. Physical relation and expression: Joint attention for human-robot interaction. *IEEE Transactions on Industrial Electronics*, vol. 50, no. 4, pp. 636–643, 2003.

[16] Y. Nagai, M. Asada, K. Hosoda. Learning for joint attention helped by functional development. *Advanced Robotics*, vol. 20, no. 10, pp. 1165–1181, 2006.

[17] H. Sumioka, K. Hosoda, Y. Yoshikawa, M. Asada. Acquisition of joint attention through natural interaction utilizing motion cues. *Advanced Robotics*, vol. 21, no. 9, pp. 983–999, 2007.

[18] M. Hashimoto, H. Kondo, Y. Tamatsu. Gaze guidance using a facial expression robot. *Advanced Robotics*, vol. 23, no. 14, pp. 1831–1848, 2009.

[19] J. K. Chu, R. H. Li, Q. Y. Li, H. Q. Wang. A visual attention model for robot object tracking. *International Journal of Automation and Computing*, vol. 7, no. 1, pp. 39–46, 2010.

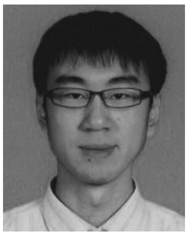
[20] C. Breazeal. Robot in society: Friend or appliance?. In *Proceedings of Autonomous Agents Workshop on Emotion-based Agent Architectures*, IEEE, Seattle, USA, pp. 11–37, 1999.

[21] Z. Leon. Robotic yo-yo: Modelling and control strategies. *Robotica*, vol. 24, no. 2, pp. 211–220, 2006.

[22] T. Petrivc, A. Gams, A. J. Ijspeert, L. vZlajpah. On-line frequency adaptation and movement imitation for rhythmic robotic tasks. *The International Journal of Robotics Research*, vol. 30, no. 14, pp. 1775–1788, 2011.

[23] A. Cooka, E. Pedro, A. kim. Robots: Assistive technologies for play, learning and cognitive development. *Technology and Disability*, vol. 22, no. 3, pp. 127–145, 2010.

[24] W. G. Song. *Robotics: Kinematics, Dynamics and Control*, Beijing, China: Science Press, pp. 18–56, 2007. (in Chinese)



Wei Wang received the M. Sc. degree in control theory and control engineering from Jiangnan University, China in 2008, and the Ph. D. degree from University of Science and Technology Beijing, China in 2012. Since 2012, he has been a faculty member at Hebei University of Engineering, China. He has published about 40 refereed journal and conference papers.

His research interests include human-robot cooperation and implicit interaction.

E-mail: wangwei83@hebeu.edu.cn (Corresponding author)
ORCID iD: 0000-0001-8876-9364

Xiao-Dan Huang received the M. Sc. degree in electronic information from University of Science and Technology Beijing, China in 2012. Since 2012, she has been a faculty member at Hebei University of Engineering, China. She has published about 10 refereed journal and conference papers.

Her research interests include robotics and implicit interaction.

E-mail: hxd10243005@sina.com