

RESEARCH

Open Access

A role of multi-modal rhythms in physical interaction and cooperation

Kenta Yonekura¹, Chyon Hae Kim^{2*}, Kazuhiro Nakadai², Hiroshi Tsujino² and Shigeki Sugano³

Abstract

As fundamental research for human-robot interaction, this paper addresses the rhythmic reference of a human while turning a rope with another human. We hypothesized that when interpreting rhythm cues to make a rhythm reference, humans will use auditory and force rhythms more than visual ones. We examined 21-23 years old test subjects. We masked perception of each test subject using 3 kinds of masks, an eye-mask, headphones, and a force mask. The force mask is composed of a robot arm and a remote controller. These instruments allow a test subject to turn a rope without feeling force from the rope. In the first experiment, each test subject interacted with an operator that turned a rope with a constant rhythm. 8 experiments were conducted for each test subject that wore combinations of masks. We measured the angular velocity of force between a test subject/the operator and a rope. We calculated error between the angular velocities of the force directions, and validated the error. In the second experiment, two test subjects interacted with each other. 1.6 - 2.4 Hz auditory rhythm was presented from headphones so as to inform target turning frequency. Addition to the auditory rhythm, the test subjects wore eye-masks. The first experiment showed that visual rhythm has little influence on rope-turning cooperation between humans. The second experiment provided firmer evidence for the same hypothesis because humans neglected their visual rhythms.

1 Introduction

In physical rhythmic human-robot interaction, rhythms provide important cues to both humans and machines. When humans operate an apparatus or control their body, they often use multi-modal rhythm perception in following their sense of internal rhythm (rhythm reference). Here, multi-modal rhythm perception means perception for independent rhythms from independent sensory organs. On the other hand, rhythm reference means single rhythm. In the operation or control, humans must make a rhythm reference from several perceptual rhythms. However, the mechanism is still not well understood.

Historically, researchers in robotics have been interested in applying the concept of human rhythm to robots for many years. In their early work on musical robots, Sugano et al. described a humanoid robot, Wabot-2, that is able to play a piano by manipulating its arms and fingers according to visually obtained music scores using its own camera [1]. Likewise, Sony

exhibited a singing and dancing robot called QRIO. Nakazawa et al. reported that HRP-2 is able to imitate the complex spatial trajectories of a Japanese traditional folk dance by using a motion capture system [2]. Shibuya et al. developed a violinist robot to realize musical expressions [3,4]. Although these robots play musical instruments, dance or sing, they were programmed in advance. Thus they had difficulties cooperating and interacting with humans.

Some researchers have more specifically examined human-robot interaction. For example, Kotosaka and Schaal [5] developed a robot that is able to play drum sessions along with a human drummer. Similarly, Michalowski et al. developed a small robot called Keepon which can move its body quickly according to musical beats [6]. Yoshii et al. developed real-time beat tracking for a robot [7]. Murata et al. extended their work to quick adaptation for changing tempo, and demonstrated its stamps, scats, and singing according to detected musical beats [8]. Later, Mizumoto et al. applied Murata's method to a Thereminist robot [9]. Hoffman and Weinberg demonstrated real-time musical sessions between a human player and a MIDI-controlled

* Correspondence: tenkai@jp.honda-ri.com

²Honda Research Institute Japan Co., Ltd., Saitama, Japan

Full list of author information is available at the end of the article

percussionist robot [10]. In their studies, robots were able to detect musical beats using auditory functions. Moreover, some other robots perform higher level interaction. Kosuge et al. described a robot dancer, MS DanceR that could perform socially dance with a human partner using just force rhythm [11]. Gentry and Murray-Smith tried a psychological human-robot-interaction research using a haptic dance leading robot [12]. Kasuga and Hashimoto demonstrated handshaking with a human [13].

Takanishi et al. developed anthropomorphic flutist robots that have lungs to send air to a flute. Takanishi's robots are able to collaborate with human players in real time [14,15].

These robots demonstrated excellent human-robot interaction, and showed that the key information from these stimuli was tempo-related, such as beat, tempo, and rhythm. Indeed, in the domain of music information processing, such tempo information is considered an essential factor for interactive systems. Dannenberg showed the world's first autonomous musical performance [16], and Vercoe and Puckette developed an automated system that adapts to human auditory rhythms [17]. Similarly, Paradiso and Sparacino developed the "Light Stick" system, which synchronizes a musical rhythm to stick motions made by a human performer [18].

However, both in robotics and music information processing, such temporal information has primarily been used as a cue by which construct robot and software applications. The utility of temporal information for executing interactive and cooperative tasks, and its relationship with various modalities have not been sufficiently examined. In human-human cooperation, one can perceive multi-modal rhythms including visual, auditory and force rhythms. For example, humans can feel force rhythm from a partner or objects operated by a partner. Likewise, human can also transmit rhythm using voice or visual motions. In studying effective rhythm cues within multi-modal rhythms, we hypothesized that when interpreting rhythm cues to make a rhythm reference, humans will use auditory and force rhythms more than visual ones. In psychological studies, evidence exists that human temporal resolutions for auditory and tactile rhythms are finer than that for visual rhythm [19]. Therefore, it is likely that humans primarily incorporate auditory and force rhythms to the neglect of visual rhythms in physical interaction.

In this article, we examine this hypothesis the use of rope turning experiments (Figure 1). Rope turning tasks are useful in exploring rhythmic physical human-robot interaction, because of their relative simplicity compared to other complex methodologies such as dancing [11]. In these tasks, experimenters are able to measure the physical rhythm of a human and a robot easily and

clearly. Moreover, both the human and the robot remain safe throughout the experiments.

2 Method

We conducted two experiments. The first experiment compares the importance of multimodal rhythms while rope-turning interaction. The second experiment confirms the amount of visual rhythm affection in various interaction conditions. In the first experiment, the sample included six participants, four males and two females, in the age range 21-23 years. The second experiment utilized two males, 22 and 23 years.

2.1 Equipment

Our equipment used for the study included a rope with a handle at each end, an eye-mask, a pair of headphones, a robot, a remote controller, a motion capture system, and a computer.

2.1.1 Rope and Handle

We used 5 m long vinyl rope weighting 44 g, a spring constant of $2.10 \times 10^2 \text{kg} \cdot \text{s}^{-2}$. We equipped each handle with a 6-axis force sensor at the ends of the rope, which is able to detect the force and moment between the handle and rope at 100 Hz sampling frequency. To reduce the force noise when a rope was untwisted, we connected each handle using an infinitely rotating mechanism. In this case, role direction moment information of the 6-axis force sensor is useless, but yaw and pitch direction one.

2.1.2 Robot Arm

We used a robot arm that was attached to a robot developed by Honda Research Institute Japan. The robot has three DoFs for the neck, three DoFs for the waist, seven DoFs for each arm, and six DoFs for each hand. The robot is equipped with two cameras, a laser range finder, a singing voice synthesizer and a speaker. Table 1 shows the specification of the arm.

2.1.3 Remote Controller

We used a Wii^a remote controller and a Wii motion plus.

2.1.4 Computer

We used a computer to control the robot arm, and capture data from the rope handles and robot.

2.1.5 Motion capture system

We used a motion capture system, VICON, with 100 Hz sampling rate to measure the position of the handles. The obtained position data was used to calculate energy transmission between the handle and the rope.

2.1.6 Force mask system

We developed a force mask system using some of these apparatuses (Figure 2). This system allows a participant to turn a rope without feeling force from the rope. When the participant turns the Wii remote, the system samples its yaw and pitch direction angular velocities by 100 Hz frequency. Then, the phase of hand direction θ

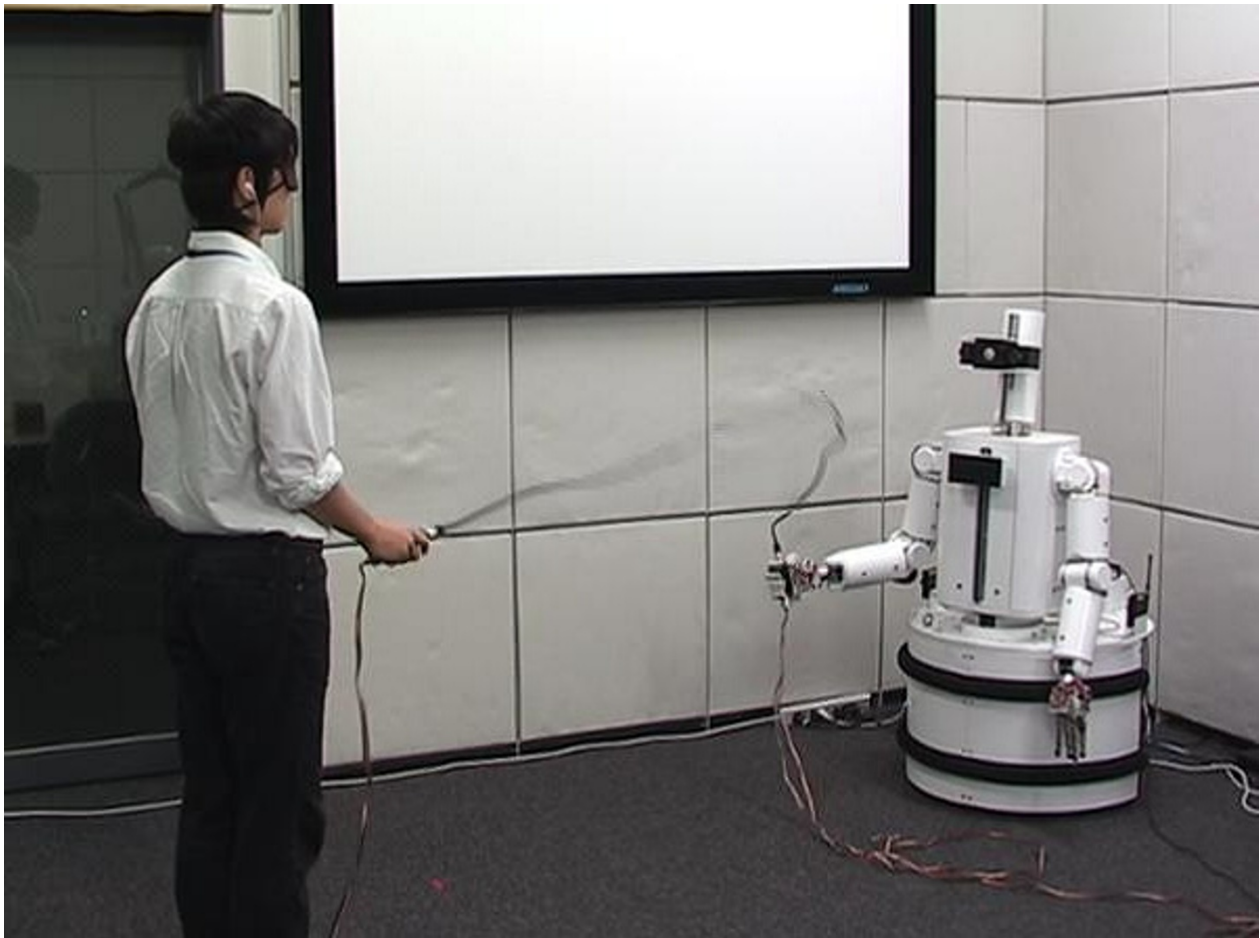


Figure 1 Rope turning experiments. In this experiment, a participant turned a rope with a robot. The participant wore eye-mask and headphones so as to inhibit his perception.

was calculated from the angular velocities based on an assumption that a hand moves on a circle. A computer sends the target position of the end-effector to the robot arm that utilizes a rope-turning algorithm [20]. We set the target position of the algorithm $T (T_x, T_y)$ equal to $(r\cos\theta, r\sin\theta)$, where r is a constant radius 0.10 m. All of the above calculations were done by using a computer, Dell Vostro 3700, that has Core i7-720QM and 8 GB memory.

Table 1 Specifications of Robot arm

Joint		Min [deg]	Max [deg]	Max speed [deg/s]
Shoulder	Revolution	-45	135	75
	Turning	0	120	113
Upper arm	Revolution	-90	135	340
Elbow	Turning	0	120	93
Front arm	Revolution	-90	90	250
Wrist	Roll	-180	180	280
	Pitch	-80	80	280

2.2 Procedures

We developed separate procedures for the two experiments. In the first experiment, a participant and an experimenter turned a rope. In the second experiment, two participants turned a rope.

2.2.1 Procedure 1

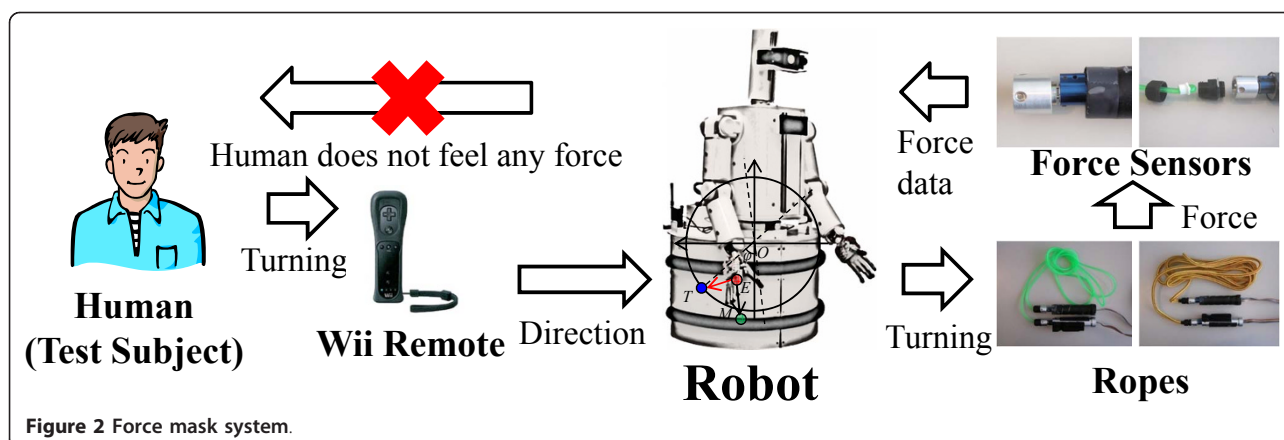
Each volunteer participant was provided informed consent prior to participation. The experimental procedure included a practice phase, followed by an instruction phase and then an experiment phase.

Practice phase Each participant turned the rope with an operator without an eye-mask or headphones. In addition, each participant used the force mask to practice controlling the rope. We continued the practice until the participant said sufficient.

Instruction phase In this following phase, we provided instructions as follows:

“We will try eight experiments.”

“Each experiment will continue for two minutes.”



“Please, turn the rope with the operator.”

“In the last four experiments, we will use the robot arm and the remote controller.”

Note that what we pronounced “experiments” is “tests” in the followings document of this paper.

Experiment phase Table 2 illustrates the combinations of masks for the tests 1 through 8. When using the force mask, the participant did not touch the rope with his/her hand. Instead of the participant, the robot arm controlled one of the handles. This phase is initiated tapping the participant’s shoulder, since each participant wore a combination of masks and the participant may therefore have been unable to know the start of the test. The operator was instructed to turn the rope with a constant rhythm almost 2 Hz, while listening to 2 Hz of sound using headphones. We measured force rhythm from the handles attached to the rope. We additionally measured the position sequence of the handles using the motion capture system. At the end of each test, the participant was tapped on the shoulder again to indicate completion.

2.2.2 Procedure 2

As with Procedure 1, each participant was provided informed consent prior to participation. We proceeded with this procedure in the order of introduction phase, followed by a practice phase and then an experimental phase.

Table 2 Combination of masks in Experiment 1

Test number	Eye mask	Headphones	Force mask
1	Off	Off	Off
2	On	Off	Off
3	Off	On	Off
4	On	On	Off
5	Off	Off	On
6	On	Off	On
7	Off	On	On
8	On	On	On

Introduction phase In this phase, the number and timing of the tests were explained to each subject, and subjects were instructed to use headphones to listen to auditory rhythms during the tests. The task of the participants was to tune the rope-turning frequency to the auditory rhythm inputted from the headphones or to that of another participant’s (while participant’s headphones did not provide rhythm). During this phase, participants were asked not to communicate through voice or gesture in any way other than their rope-turning motion.

Practice phase In this phase, participants practiced a set of tests without the use of eye masks. Headphones provide the same rhythm as in following experiment phase. After each test, participants rested to prevent excessive arm fatigue.

Experiment phase In this phase, participants attempted three sets of tests. We show the combination of eye-masks that the participants put during the tests in Table 3.

3 Results

We show the results of Experiments 1 and 2 that consist of Procedure 1 and 2, respectively.

3.1 Results of Experiment 1

After the experiment, we validated error E between the participant and operator using the following equation.

$$E(t) = \|\dot{\theta}_p(t) - \dot{\theta}_o(t)\| \quad (1)$$

$\dot{\theta}_p$ and $\dot{\theta}_o$ are angular velocities of the handles on both the participant and operator sides. Angular velocities

Table 3 Combination of eye-masks in Experiment 2

Test number	Participant 1	Participant 2
1	Off	Off
2	On	Off
3	On	On

were calculated via the rope-turning frequency by detecting the peaks of force direction data obtained from the rope handles.

When error is zero, both operator and participant are successfully cooperating. Time average of 5,000 error data for a male participant is shown in Figure 3. *T*-tests indicated that the differences between any pair of experiments in Figure 3 were significant at $p \leq .05$, except between experiments three and six.

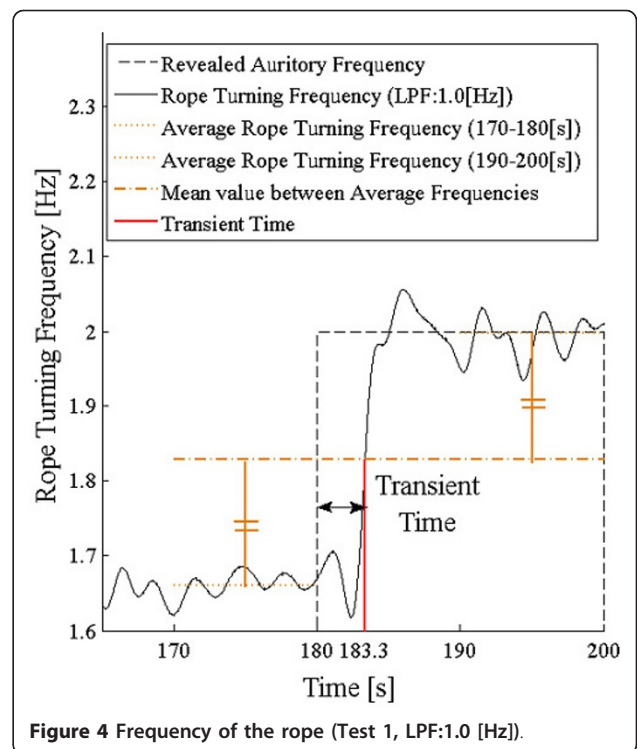
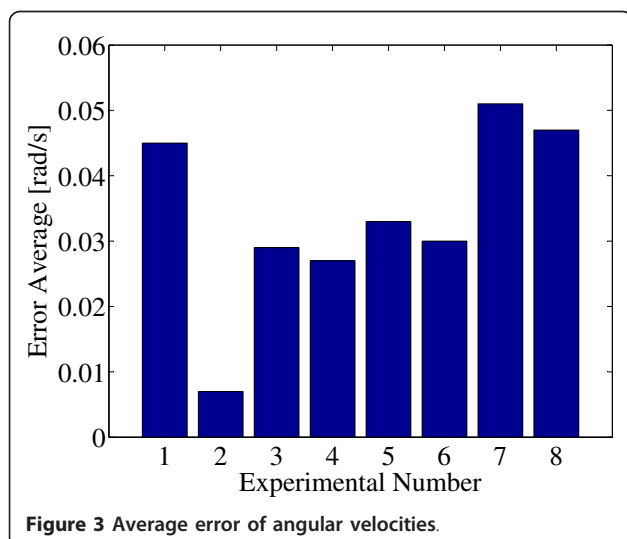
In Test 1, without any mask, the error of the participant was about 0.045 rad/s. When the participant used a mask (Tests 2, 3, and 5), the error decreased. Error tended to increase with increased mask use.

3.2 Results of Experiment 2

We analyzed the rotation frequency of the rope handles based on handle's angular velocities $\dot{\theta}_{p1}$, $\dot{\theta}_{p2}$, and a rope turning angular velocity $(\dot{\theta}_{p1} + \dot{\theta}_{p2})/2$.

3.2.1 Test 1

Figures 4 and 5 show the rope's temporal frequency while the two participants were turning it. We applied low-path filters using cutoff frequencies of 0.5 and 0.1 Hz to raw data, respectively. Table 4 shows maximum, average, and minimum frequencies in addition to average error between the handle's turning frequencies, and shows the presented auditory frequencies during those respective time spans. The average error refers average of absolute difference between presented auditory frequency and rope-turning frequency. Table 5 shows the amount of time between the moment the presented auditory rhythm was switched and the moment the rope-turning frequency crossed the mean of the pre and post frequencies. The schematic of this calculation is illustrated in Figure 6. Prior to frequency calculation, we used a low path filter with a cutoff frequency 0.1 Hz in



these cases. This figure shows transient time at 183.3 s for example. This time is mean of the pre and post frequencies that are calculated in their respective time periods (see, Tables 4, 6, and 7).

3.2.2 Tests 2 and 3

As was done for Test 1 results, Figures 7, 8, and Table 6 show results from Test 2. Table 5 shows the amount of time required for transition.

Finally, Figures 9, 10, and Table 7 show the results of Test 3. Again, Table 5 shows the amount of time required for transition.

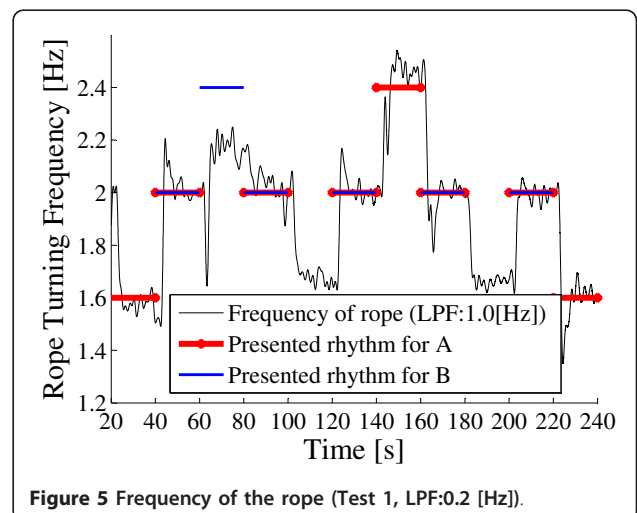


Table 4 Results of Test 1

Period	1	2	3	4	5	6	7	8	9	10	11	12	13
Time s	0-20	20-40	40-60	60-80	80-100	100-120	120-140	140-160	160-180	180-200	200-220	220-240	240-260
Presented rhythm for A Hz	2.0	1.6	2.0	-	2.0	-	2.0	2.4	2.0	-	2.0	1.6	2.0
Presented rhythm for B Hz	2.0	-	2.0	2.4	2.0	-	2.0	-	2.0	-	2.0	-	2.0
Maximum frequency Hz	neglect	1.73	2.25	2.28	2.17	1.74	2.18	2.60	2.15	1.75	2.08	1.73	neglect
Average frequency Hz	neglect	1.58	2.03	2.17	2.02	1.67	2.02	2.44	1.99	1.66	2.00	1.58	neglect
Minimum frequency Hz	neglect	1.44	1.93	2.06	1.83	1.58	1.91	2.07	1.83	1.57	1.90	1.23	neglect
Average error Hz	neglect	0.03	0.00	0.23	0.01	-	0.01	0.07	0.02	-	0.00	0.02	neglect

Table 5 Required time for transition

Period	2-3	3-4	4-5	5-6	6-7	7-8	8-9	9-10	10-11	11-12
Interval s (Test 1)	3.44	4.67	3.00	3.13	3.12	3.39	2.68	2.53	3.37	3.14
Interval s (Test 2)	1.35	2.69	0.99	0.79	1.62	3.14	1.33	1.17	1.24	1.66
Interval s (Test 3)	1.42	3.63	1.02	0.00	2.00	1.72	0.91	1.13	0.00	3.02

4 Discussion

4.1 Hypothesis

Results of Experiment 1 support our hypothesis that when interpreting rhythm cues to make a rhythm reference, humans will use auditory and force rhythms more than visual ones. In Experiment 1, the error of Test 1 was very large. This might have been a result of insufficient practice. Except for Test 1, the participant's error increased when using a larger number of masks. If we were to ignore the first test, our results supports our hypothesis because there are only small error differences between 'on' and 'off' for the visual mask (see differences between Tests 3 and 4, Tests 5 and 6, and Tests 7 and 8).

Similarly, results from Experiment 2 provide firm evidence for our hypothesis. For example, Test 3 shows very small average errors, and these errors are almost the same as those in Test 1.

This strongly shows that both participants cooperated without the use of visual rhythms. In other words, visual rhythm was almost not required to cooperate in this rope-turning task. This result also suggests that human may rely on modalities that have higher perceived resolutions. Further research would necessary to confirm this.

4.2 Practice for the task

Our findings underscored the difficulty in examining the performance of non-practiced participants. For example, in Experiment 1, we did not provide sufficient practice time (only the practice time to use the robot arm). Therefore, the results suggested that participants were able to quickly adapt to the task. In Experiment 2, we attempted not to collect data from non-practiced participants by letting the participants practice sufficiently. Subsequently, there was little difference between the early period (Test 1, Table 4) and last period of the experiments (Test 3, Table 7). From these results, we believe that there was little practice while conducting this experiment. Therefore, collecting data from non-practised participants might be so difficult, since there is little time until the completion of the practice.

4.3 Eye-mask provided slightly better results

In Experiment 2, transition time (Table 5) and average error (Tables 4, 6 and 7) show slightly better results in Test 3. There are two possible explanations for this finding. The first possibility is participant's practice. Though we set a long practice time in this experiment, the participants may have continued their sense of practicing incrementally throughout the duration of Tests 1 through 3. The second possibility is the effect of eye masks. Eye masks may have enhanced participants

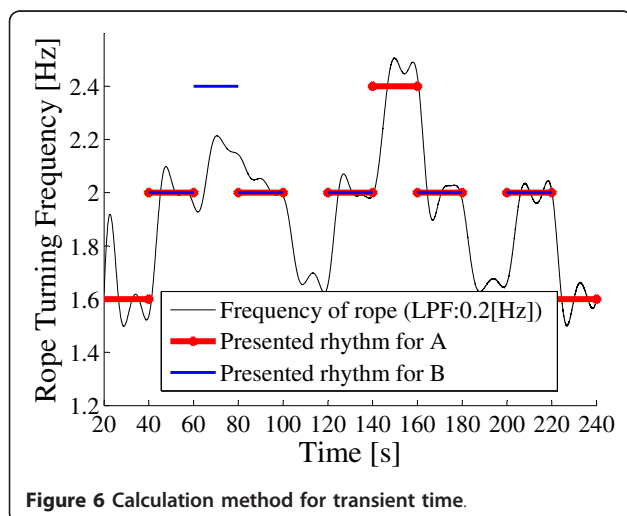


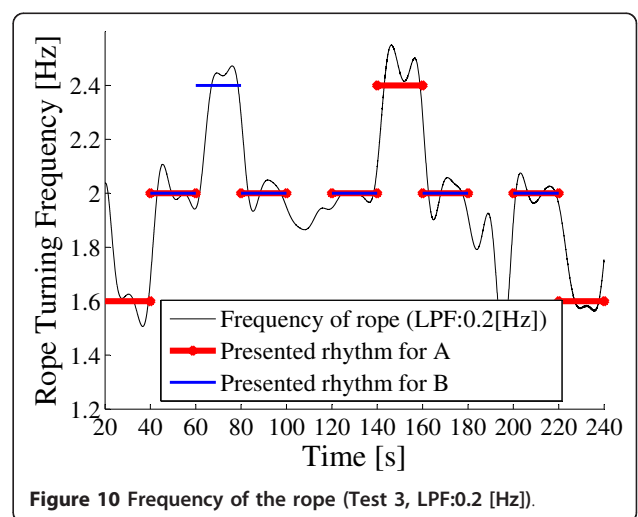
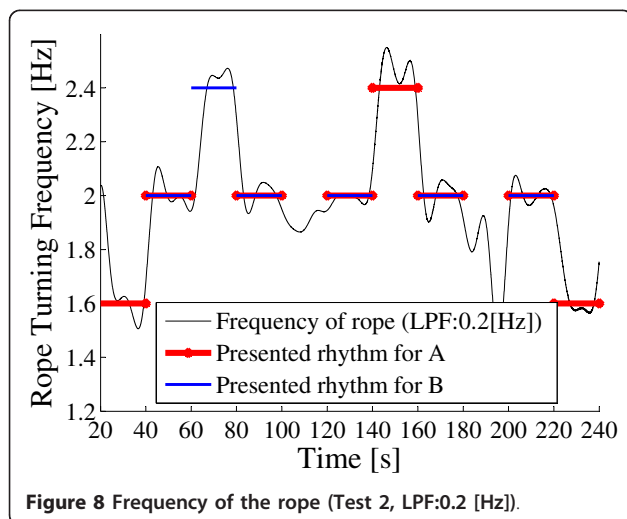
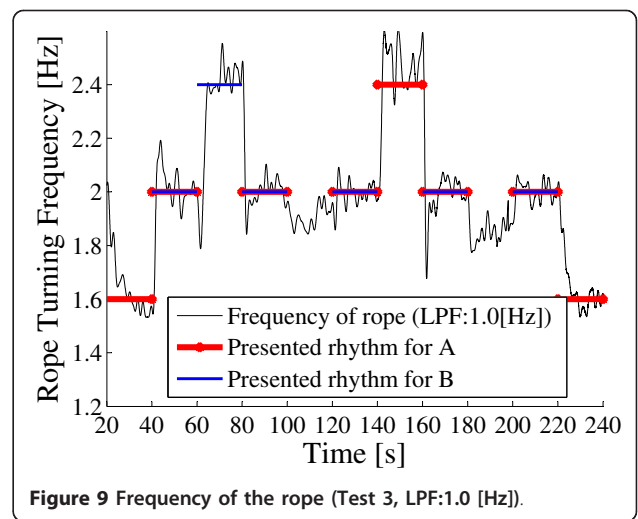
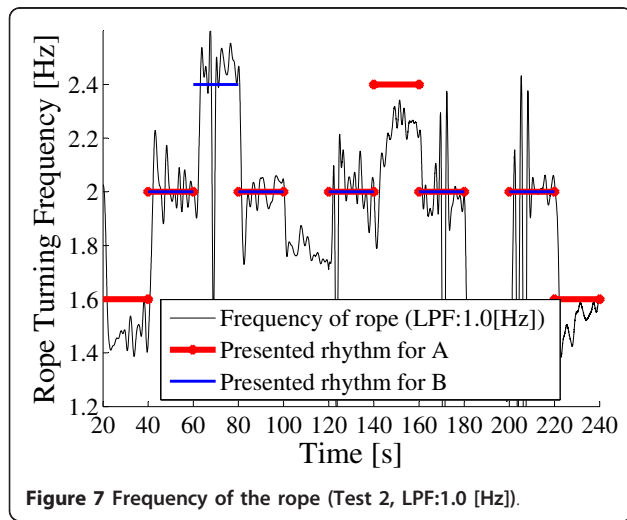
Figure 6 Calculation method for transient time.

Table 6 Results of Test 2

Period	1	2	3	4	5	6	7	8	9	10	11	12	13
Time s	0-20	20-40	40-60	60-80	80-100	100-120	120-140	140-160	160-180	180-200	200-220	220-240	240-260
Presented rhythm for A Hz	2.0	1.6	2.0	-	2.0	-	2.0	2.4	2.0	-	2.0	1.6	2.0
Presented rhythm for B Hz	2.0	-	2.0	2.4	2.0	-	2.0	-	2.0	-	2.0	-	2.0
Maximum frequency Hz	neglect	1.68	2.26	2.60	2.10	1.92	2.17	2.39	2.16	1.60	2.12	1.64	neglect
Average frequency Hz	neglect	1.47	2.02	2.46	1.98	1.78	2.02	2.24	1.97	1.32	2.01	1.52	neglect
Minimum frequency Hz	neglect	1.26	1.80	2.31	1.63	1.65	1.88	2.06	1.81	1.15	1.90	1.34	neglect
Average error Hz	neglect	0.12	0.01	0.03	0.01	-	0.01	0.14	0.14	-	0.01	0.04	neglect

Table 7 Results of Test 3

Period	1	2	3	4	5	6	7	8	9	10	11	12	13
Time s	0-20	20-40	40-60	60-80	80-100	100-120	120-140	140-160	160-180	180-200	200-220	220-240	240-260
Presented rhythm for A Hz	2.0	1.6	2.0	-	2.0	-	2.0	2.4	2.0	-	2.0	1.6	2.0
Presented rhythm for B Hz	2.0	-	2.0	2.4	2.0	-	2.0	-	2.0	-	2.0	-	2.0
Maximum frequency Hz	neglect	1.73	2.14	2.58	2.16	2.02	2.08	2.68	2.12	2.11	2.19	1.81	neglect
Average frequency Hz	neglect	1.60	2.01	2.43	2.01	1.90	2.00	2.48	2.00	1.85	2.00	1.60	neglect
Minimum frequency Hz	neglect	1.47	1.85	2.34	1.83	1.79	1.90	2.07	1.84	1.72	1.85	1.45	neglect
Average error Hz	neglect	0.03	0.02	0.03	0.02	-	0.01	0.04	0.02	-	0.00	0.00	neglect



abilities to concentrate on auditory and/or force rhythms by masking the less-useful visual rhythms. Another experiment would be necessary to confirm these hypotheses.

4.4 For further confirmation

Though the obtained data supports our hypothesis strongly, the relationship to visual temporal perception characteristic [19] is still weak. To generalize our findings to many kinds of interaction, we need to confirm the relationship by improving our methodology.

In our experiments, completed individual difference elimination was difficult, because the experiments required a large scale system and the available term of the system was limited. We hope that further researches will be done to get conclusion about the difference.

5 Conclusion

We conducted two experiments to confirm the hypothesis that when interpreting rhythm cues to make a rhythm reference, humans will use auditory and force rhythms more than visual ones. The first experiment showed that visual rhythm has little influence on rope-turning cooperation between humans. The second experiment provided firmer evidence for the same hypothesis because humans neglected their visual rhythms. Further research with other types of tasks (for, e.g., cooperative carrying task, dancing task, and so on) is needed generalize this finding.

Endnote

^aNintendo.

Acknowledgements

We would like to thank Dr. Kazuhito Yokoi, AIST, Japan for his help. We would like thank Mr. Hideki Kenmochi and Mr. Yasuo Yoshioka, YAMAHA Corporation for their help. We also would like thank, Mr. Satoshi Miura, Waseda University, Fujie Laboratory for his help.

Author details

¹The Dept. of Intelligent Interaction Technologies at Tsukuba Univ., Ibaraki, Japan ²Honda Research Institute Japan Co., Ltd., Saitama, Japan ³School of Creative Science and Engineering at Waseda Univ., Tokyo, Japan

Competing interests

The authors declare that they have no competing interests.

Received: 8 August 2011 Accepted: 2 March 2012

Published: 2 March 2012

References

1. Science and Engineering Research Laboratory, Special Issue on Wabot-2. Bull. No. 112 (Waseda University, Tokyo, 1985), (authors unknown)
2. A Nakazawa, S Nakaoka, K Ikeuchi, K Yokoi, Imitating human dance motions through motion structure analysis. in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems* (Lausanne). **3**, 2539–2544 (2002)
3. K Shibuya, S Sugano, The effect of KANSEI information on human motion - basic model of KANSEI and analysis of human motion in violin playing. in

- Proceedings of the IEEE International Workshop on Robot and Human Communication* (Tokyo) 89–94 (1995)
4. K Shibuya, S Matsuda, A Takahara, Toward developing a violin playing robot - bowing by anthropomorphic robot arm and sound analysis. in *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication* (Jeju island) 763–768 (2007)
5. S Kotosaka, S Schaal, Synchronized robot drumming by neural oscillators. in *Proceedings of the International Symposium Adaptive Motion of Animals and Machines* (Montreal) (2000)
6. Marek Michalowski, H Kozima, S Sabanovic, A dancing robot for rhythmic social interaction. in *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction* (Washington DC) 89–96 (2007)
7. K Yoshii, K Nakadai, T Torii, Y Hasegawa, H Tsujino, K Komatani, T Ogata, HG Okuno, A biped robot that keeps steps in time with musical beats while listening to music with its own ears. in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems* (San Diego) 1743–1750 (2007)
8. K Murata, K Nakadai, K Yoshii, R Takeda, T Torii, HG Okuno, Y Hasegawa, H Tsujino, A robot singer with music recognition based on real-time beat tracking. in *Proceedings of the International Conference on Musical Information Retrieval* (Pennsylvania) 199–204 (2008)
9. T Mizumoto, H Tsujino, T Takahashi, T Ogata, HG Okuno, Thereminist robot: development of a robot theremin player with feedforward and feedback arm control based on a Theremin's Pitch model. in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems* (Saint Louis) 2297–2302 (2009)
10. G Hoffman, G Weinberg, Gesture based human-robot Jazz improvisation. in *Proceedings of the IEEE/RAS International Conference on Robots and Automation* (Anchorage) 582–587 (2010)
11. K Kosuge, T Hayashi, Y Hirata, R Tobiyama, Dance partner robot - MS DanceR. in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems* (Las Vegas). **3**, 3459–3464 (2003)
12. S Gentry, R Murray-Smith, Haptic dancing: human performance at haptic decoding with a vocabulary. in *Proceedings of IEEE International conference on Systems Man and Cybernetics* (Washington DC). **4**, 3432–3437 (2003)
13. T Kasuga, M Hashimoto, Human-robot handshaking using neural oscillators. in *Proceedings of the IEEE International Conference on Robotics and Automation* (Barcelona) 3802–3807 (2005)
14. A Takanishi, M Sonehara, H Kondo, Development of an anthropomorphic flutist robot WF-3Rll. in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems* (Osaka). **1**, 37–43 (1996)
15. J Solis, K Taniguchi, T Ninomiya, K Petersen, T Yamamoto, A Takanishi, Implementation of an auditory feedback control system on an anthropomorphic flutist robot inspired on the performance of a professional flutist. *Advanced robotics*. **23**, 1849–1871 (2009). doi:10.1163/016918609X12518783330207
16. RB Dannenberg, An on-line algorithm for real-time accompaniment. in *Proceedings of the International Computer Music Conference* (Paris) 193–198 (1985)
17. B Vercoe, M Puckette, Synthetic rehearsal, Training the synthetic performer. in *Proceedings of the International Computer Music Conference* (Paris) 275–278 (1985)
18. JA Paradiso, F Sparacino, Optical tracking for music and dance performance. in *Proceedings of the Optical 3D Measurement Techniques* (Zurich) 11–18 (1998)
19. M Lotze, M Wittmann, Daily rhythm of temporal resolution in the auditory system. *Cortex*. **35**, 89–100 (1999). doi:10.1016/S0010-9452(08)70787-1
20. CH Kim, K Yonekura, H Tsujino, S Sugano, Physical control of the rotation of a flexible object - rope turning with a humanoid robot. *Advanced robotics*. **25**, 491–506 (2011). doi:10.1163/016918610X551791

doi:10.1186/1687-4722-2012-12

Cite this article as: Yonekura et al.: A role of multi-modal rhythms in physical interaction and cooperation. *EURASIP Journal on Audio, Speech, and Music Processing* 2012 **2012**:12.