Development of an Embodied Group Entrainmacent Response System to Express Interaction-Activated Communication

Yutaka Ishii$^{(\boxtimes)}$ and Tomio Watanabe

Okayama Prefectural University, Kuboki 111, Soja, Okayama, Japan
{ishii,watanabe}@cse.oka-pu.ac.jp

Abstract. In situations that demand audience participation, such as lectures and speeches, the sense of unity of the communication place is created by mutual interactions between the performer and audience. Therefore, the speaker's willingness to talk increases with the increase in the responses of the listeners, thus strengthening the interaction. We have developed a speech-driven embodied entrainment toy robot called Pekoppa that generates communicative motions and actions such as nods for entrained interactions from speech rhythm based only on voice input. In this research, we developed a system that activates the interaction by changing the embodied response by nodding from multiple toy robots based on speech inputs. Moreover, we confirmed the effectiveness of the system by sensory evaluation and behavior analysis through experiments on 24 participants.

Keywords: Embodied communication · Group entrainment · Audience characters

1 Introduction

In scenes with audience participation, such as lectures and speeches, the sense of unity of the communication place depends on the mutual interactions between the performer and audience. Visualizing the excitement of the audience is considered important for effective communication. Sejima has defined the state in which embodied interactions, such as body motions and voice responses, are activated as "interaction-activated communication", and the effectiveness of this estimation model has been confirmed by the heat conduction equation using speakers' voice inputs [1].

In contrast, group communication is formed by a gradual gathering of people around the speaker that causes excitement. Therefore, the speaker's willingness to talk depends on the number of listeners, who gradually form a group and activate the interaction.

In this research, we developed a system that activates interaction through changes in the embodied response by nodding using multiple objects based on speech inputs. We confirmed the effectiveness of the system by evaluation experiments.

2   Related Work

Watanabe et al. developed a speech-driven embodied interactive actor called Inter-Actor, with functions of both speaker and listener, for activating human interaction and communication by generating expressive actions and motions that are coherently related to speech inputs [2]. Giannopulu et al. has reported that minimalistic artificial environments, such as toy robots, could be considered as the root of neuronal organization and reorganization with the potential to improve brain activity in children with autism [3]. Moreover, they analyzed nonverbal and verbal information associated with the heart rate and emotional feeling in ASD and neurotypical children respectively. As a result, analogies of heart rate between ASD and neurotypical children were expressed when the human was the 'passive' actor and the robot was the 'active' actor; disanalogies were observed when the human was the 'active' actor [4].

Regarding the collective communication research by robot, Karatas et al. proposed driving agents called NAMIDA (Navigational Multiparty based Intelligent Driving Agents) as three friendly interfaces those sit on the dashboard inside a car [5]. Rubenstein et al. proposed an open-source, low cost robot called Kilobot that designed to make testing collective algorithms on hundreds or thousands of robots [6].

3   Overview of an Embodied Group Entrainment Response System

3.1   Concept

In a one-to-many dialogue situation such as a speech or an over-the-counter sale, the presence of the audience greatly affects the activation of the interaction. Active involvement of audiences influences the interaction with other audiences and motivates the speaker's utterances. In this research, we propose a communication support system showing activation of interaction by an increase in the audience objects. A communication effect is obtained by group entrainment using a response model based only on the speech input (Fig. 1).
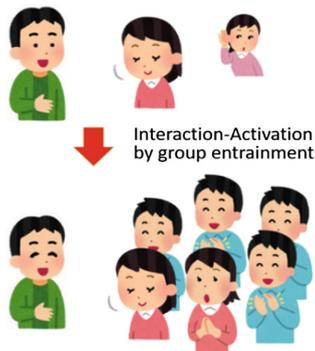

Fig. 1. Concept.

3.2   Interaction Model

A listener's interaction model includes a nodding reaction model that estimates the nodding timing from a speech ON-OFF pattern and a body reaction model linked to the nodding reaction model (Fig. 2). A hierarchy model consisting of two stages, macro and micro, predicts the timing of the nodding. The macro stage estimates whether a nodding response exists in a duration unit that consists of a talkspurt episode $T(i)$ and the subsequent silence episode $S(i)$ with a hangover value of 4/30 s. The estimator $M_u(i)$ is a moving-average (MA) model, expressed as the weighted sum of unit speech activity R (i) in (1) and (2). When $M_u(i)$ exceeds the threshold value, the nodding $M(i)$ is also an MA model, estimated as the weighted sum of the binary speech signal $V(i)$ in (3). The body movements are related to the speech input at a timing over the body threshold. The body threshold is set lower than that of the nodding prediction of the MA model that is expressed as the weighted sum of the binary speech signal to nodding.

$$M_u(i) = \sum_{j=1}^{J} a(j)R(i - j) + u(i) \qquad (1)$$

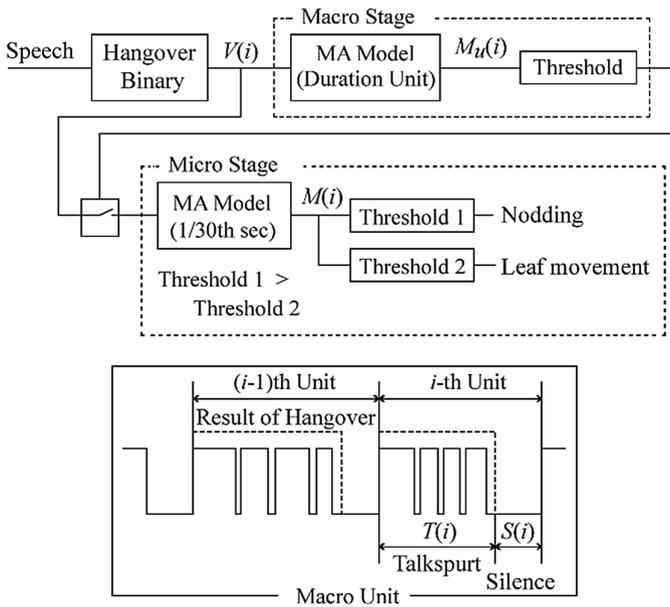$$R(i) = \frac{T(i)}{T(i) + S(i)} \qquad (2)$$

Fig. 2. Listener's interaction model.

$a(j)$: linear prediction coefficient
$T(i)$: talkspurt duration in the i-th duration unit
$S(i)$: silence duration in the i-th duration unit
$u(i)$: noise

$$M(i) = \sum_{k=1}^{K} b(j)V(i-j) + w(i) \tag{3}$$

$b(j)$: linear prediction coefficient
$V(i)$: voice
$w(i)$: noise

3.3   Development of System Prototype

System Prototype Using LED. In this research, we developed a prototype system based on the concept of light emission by an LED (Fig. 3). Based on the listener interaction model, the listeners group is represented by the blinking LED corresponding to the speaker's voice input. However, a few test users opined that an LED is difficult to recognize as an independent object, and it seems like a product of art on the panel (Fig. 4).
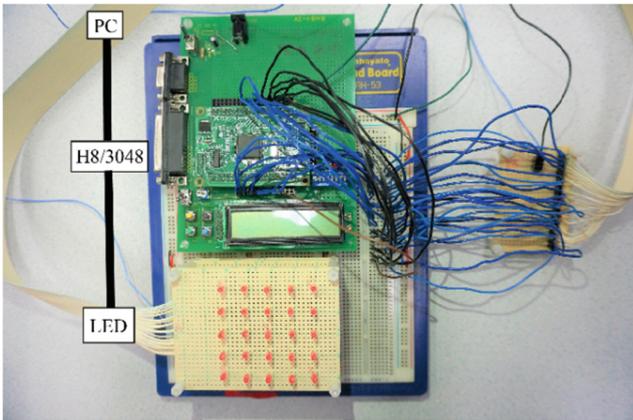

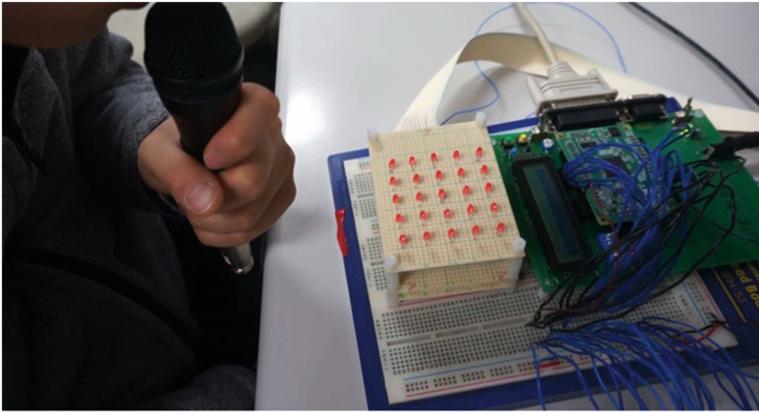**Fig. 3.** System prototype using LED.

Fig. 4. Example of a using scene.

System Construction Using Bilobed Plant Toys. To express the increase in the number of independent objects reacting to the speaker, we assumed a situation in which toy plants that perform large interaction actions against human speech are spread. Based on a model predicting nodding from the voice of the interlocutor, 25 of the bilobed plant toys (Fig. 5 Pekoppa: SegaToys 2008) that perform nodding reactions automatically are placed on a 700 mm square plate. They are arranged in 5 rows on the board and express the group (Fig. 6). Vocal utterance is captured by the microphone input on the PC and the timing estimation result of the nodding start is transmitted from the PC to the H8/3048 micro-computer by serial communication. The H8 microcomputer individually controls the behavior of the plant-type toy, so that any toy can nod and react at the timing of the nod starting. Interaction activation based on population entrainment can thus be considered by expressing increase of listener individual freely.
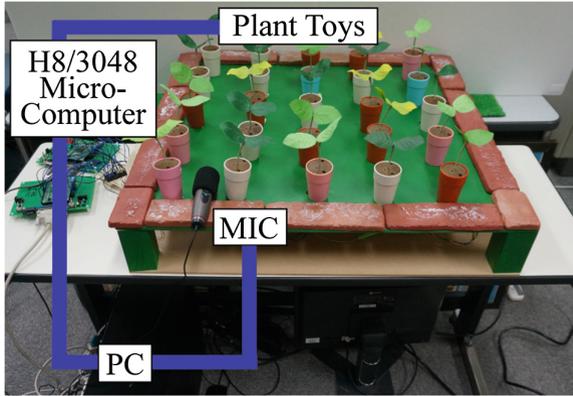

Fig. 5. Pekoppa: a bilobed plant toy.

Fig. 6. System prototype using bilobed plant toys.

4 System Evaluation Experiment

4.1 Experimental Setup

We conducted an evaluation experiment to examine the motivation of the speaker's utterance. Experiments were conducted in a three-mode comparison. These are Mode A in which all the plant type toys nod from the beginning of the nodding timing, Mode B in which the plant type toys nod in a row by side frequency (Fig. 7), and Mode C in which the number of plant type toys nods increases from one to semicircular. (Figure 8). Each participant was presented with the three modes in a random order to eliminate any ordering effect. Experiment participants were 24 male/female students aged 18 to 24 years.
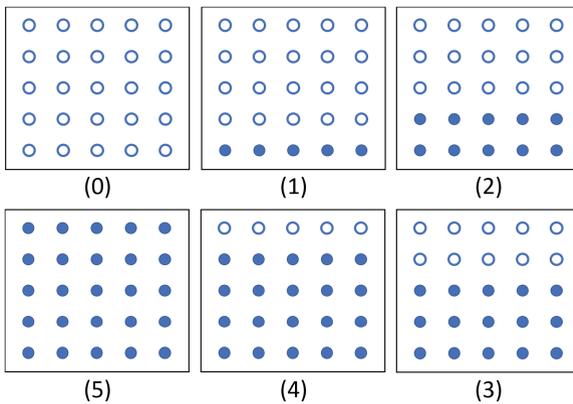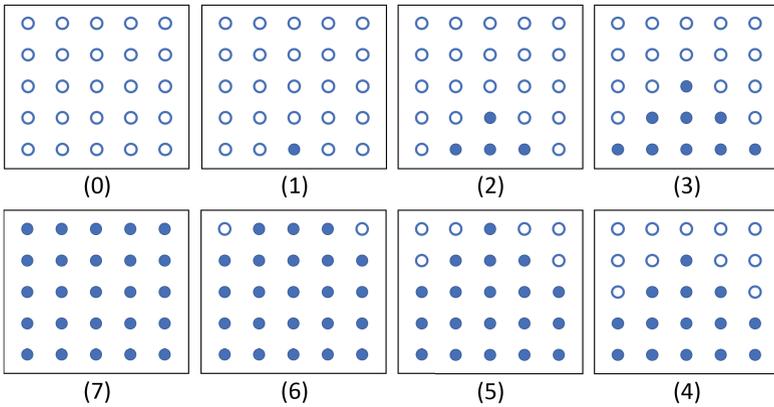

Fig. 7. Increasing pattern in Mode B.

Fig. 8. Increasing pattern in Mode C.

At first, the participants were introduced to the three operational modes and the differences between them while using the system. Next, the subjects were instructed to perform a pairwise comparison of each mode for an overall evaluation. Since three comparisons were required, the experiment was conducted three (= 3C2) times. Then, the questionnaire was examined using a-3 (not at all) to 3 (extremely) bipolar rating scale. The subjects evaluated the three modes in terms of six items—preference, enjoyment, ease of talking, comfortableness, interaction-activation, and usability. Finally, we conducted a free utterance experiment to stop the utterance when the participants thought that it was sufficient. The upper limit of speech time was set to 300 s.

4.2 Result

The results of the paired comparison for the three modes are shown in Table 1. Figure 9 shows the calculated results of the evaluation provided in Table 1, based on the Bradley–Terry model given in Eq. (4). Mode C in which the number of plant-type toys' nods increase from one to semicircular was evaluated most affirmatively, with Mode A and Mode B following the order.

Table 1. Result of the paired comparison.

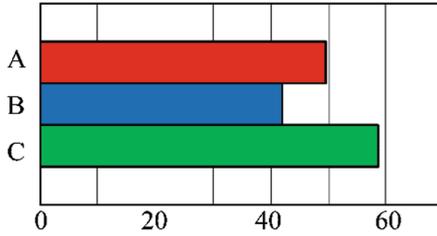|	A	B	C	Total
A		13	11	24
B	11		10	21
C	13	14		27

Fig. 9. Preference strength π for each mode.

$$P_{ij} = \frac{\pi_i}{(\pi_i + \pi_j)}$$
$$\sum_i \pi_i = const.(= 100) \tag{4}$$

(π_i: intensity of i, P_{ij}: probability of judgement that i is better than j.)

Figure 10 shows the result of the sensory evaluation in the experiment. Significant differences between each of the three modes were obtained by administering Friedman's test. Significant differences were also obtained by administering the Wilcoxon's rank test for multiple comparisons. As a result, a significant level of 5% was obtained for the "Interaction-Activation" factors between Modes A and C. The plant type toy that responds to the speaker's voice seemed to come closer to the experiment participant, and the speaker gradually felt the interaction activation.
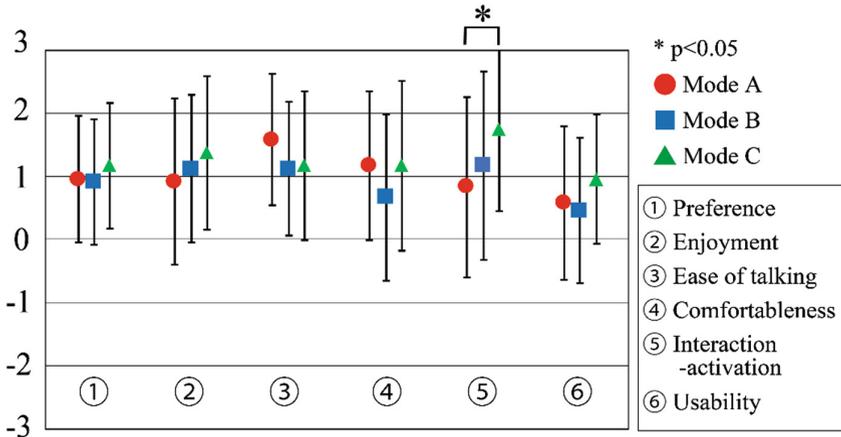

Fig. 10. Results of seven point bipolar rating.

Further, the result of the free conversation experiment was that it was spoken for a long time in Mode A and Mode C in comparison with Mode B (Fig. 11). In Mode B, in the free description section of the questionnaire to the experiment participants, there were opinions that increasing the number of mechanical and monotonous instruments felt unnatural. It may be caused by the fact that the way the audiences gathered is not

natural. From the results, it is concluded that not only does the response increase but the reaction that excites the speaker also increases, leading to the motivation from the speaker's utterance.
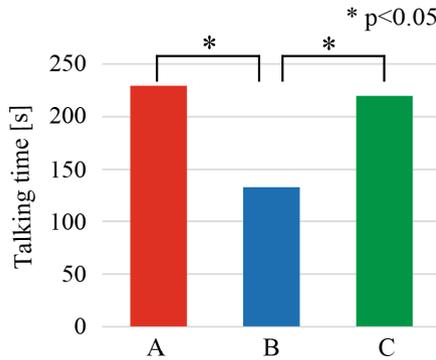

Fig. 11. Talking time of participants in each mode.

5   Conclusion

In this research, we have developed a system that presents the activation of interaction through a change in the embodied response by nodding of multiple objects based on speech input. We confirmed the effectiveness of the system by the evaluation experiment.

References

1. Sejima, Y., Watanabe, T., Jindai, M.: Estimation model of interaction-activated communication based on the heat conduction equation. Journal of Advanced Mechanical Design, Systems, and Manufacturing **10**(9), 1–11 (2016). Paper No. 15-00548
2. Watanabe, T., Okubo, M., Nakashige, M., Danbara, R.: InterActor: speech-driven embodied interactive actor. Int. J. Hum. Comput. Inter. **17**(1), 43–60 (2004)
3. Giannopulu, I., Montrrynaud, V., Watanabe, T.: Minimalistic toy robot to analyze a scenery of speaker-listener condition in autism. Cogn. Process. Springer **17**(2), 195–203 (2016)
4. Giannopulu, I., Terada, K., Watanabe, T.: Communication using robots: a perception-action scenario in moderate ASD. J. Exp. Theor. Artif. Intell. (2018). https://doi.org/10.1080/0952813x.2018.1430865
5. Karatas, N., Yoshikawa, S., De Silva, P.R.S., Okada, M.: NAMIDA: multiparty conversation based driving agents in futuristic vehicle. In: Kurosu, M. (ed.) HCI 2015. LNCS, vol. 9171, pp. 198–207. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-21006-3_20
6. Rubensteina, M., Ahlera, C., Hoffa, N., Cabrerab, A., Nagpala, R.: Kilobot: A low cost robot with scalable operations designed for collective behaviors. Robot. Auton. Syst. (2013). https://doi.org/10.1016/j.robot.2013.08.006