*Ishihara & Ishihara, Gerontechnology* (2023) 22:2
https://doi.org/10.4017/gt.2023.22.2.ish.08

# The Talking Teddy Bear for social conversation: A pilot study

Keiko Ishihara PhD[a,*], Shigekazu Ishihara PhD[a]

[a]*Hiroshima International University, Hiroshima 739-2695, Japan; *Corresponding author: k-ishiha@hirokoku-u.ac.jp*

**Abstract**

This research investigates the effectiveness of a talking teddy bear robot in facilitating creative conversations with an older adult (TN) in her 70s and living alone. She interacted with the robot for three days. The robot responded to TN's phonological information to encourage continuation of the dialogue. Analysis of TN's speech patterns showed active participation, such as discussing daily events, sharing opinions, and asking questions. TN rated the robot's responses as fair and appreciated its friendly performance. The expressive abilities of the robot both facilitated communication and improved social interactions.

**Keywords:** talking robot, response, sociality, emotions

## Introduction

In Japan, the number of older people living alone continue to rise, exceeding 590 million in 2015 (Cabinet Office 2019). Older people living alone are at risk of cognitive decline due to a lack of opportunities to talk to others. Speech production and swallowing involve common muscle groups such as the larynx, tongue, lingual jaw and lips, with close coordination between swallowing and breathing while eating or drinking (Nishio & Niimi 2004). Movements of the oral articulators influence swallowing (Inoue et al., 2014). A decline in speech-related muscle strength also affects lung and swallowing function.

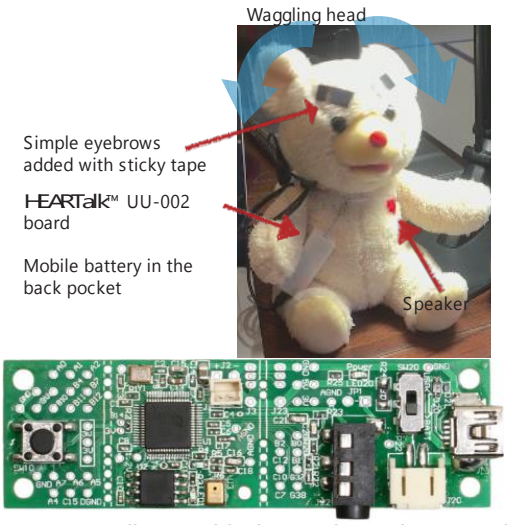


*Figure 1. Talking teddy bear robot (above) and HEARTalk™ UU-002 board (below)*

Anthea Tinker's concept of creativity in older adults, Active Aging, is not just about physical activity. She emphasizes that creativity and creative activities are important for maintaining cognitive abilities, both socially and individually (Price & Tinker 2014).

Recently, communication robots with advanced speech recognition capabilities have appeared. In some elderly care facilities in Japan, they are used to interact with users, to provide exercise guidance, and to manage medication. However, in such cases, elderly seem to be guided passively rather than creatively. If the robot doesn't have a high level of intelligence, how will the user interact with it?

The communication function of the robot devised for this study makes the robot passive. It uses the phonological information in human speech only, not its content, to encourage the user to continue speaking. In this paper we analyze how an elderly person responds to such a robot and how the conversation develops.

## Method

### Teddy bear hardware

Our aim in designing the hardware was to achieve a coordinated response that acts as a social bond between speaker and listener. The chosen teddy bear (Himitsu no Kuma-chan = Mysterious teddy bear from T.Arts Inc.) already has a built-in speaker that emits sounds while shaking its head left and right. A HEARTalk™ board (Yamaha Corporation 2017) created the audio output. It analyzes pitch and prosody of the user's voice in real time and responds with pre-recorded messages. It aims to analyze auditory emotional expressions in speech, rather

*Table 1. Recorded responses for the robot (translated from local Hiroshima dialect)*

| Function | Response of teddy bear |
|---|---|
| Approval | Sounds great! / Yes, indeed! / Nice! / Indeed! |
| Applause | That's impressive. |
| Response | I see. / Oh, I get it. / Oh, did you? / I knew it. / Yeah, I guess. / Yes. |
| Sympathy | You've had a tough day. |
| Surprise | Really? |
| Urging to go on | And then what? |
| Reaction | It's terrible! / Good for you. / How awful. |
| Asking to repeat* | Huh? / What did you say? |

*After small volume speech

than recognizing what is being said. We attached this board to the teddy bear to create a robot that responds to emotional expressions in the user's speech by answering, while shaking its head. We added eyebrows to the teddy bear to give it the appearance of a talking bear (*Figure 1*).

Tone determinants were classified into six groups of prosody (a combination of loudness, syllable length, and word endings). The pitch of the voice is recognized in 12 levels. A total of 84 types of response voice data combining 6 prosody groups and 12 pitches can be stored in the HEARTalk board and output according to the real-time analysis results of the user's speech. Instead of using 12 pitch levels, we grouped them into 3 levels and prepared a total of 19 response voices (*Table 1*).

*Table 2. Samples of the human-bear dialogue*

**Sample dialogue 1 (Day 2):**
TN told the bear that the gardener had come, pruned the plants, and the garden was now clean.

| | |
|---|---|
| TN: | So, |
| Bear: | How awful. |
| TN: | It's a tough job, though, ... |
| Bear: | Yeah, I guess. |
| TN: | But when people who love gardens look at my garden, they think, This house is well kept, isn't it? |
| TN: | There are people who think like that. |
| Bear: | Indeed! |
| TN: | Are there? |

**Sample dialogue 2 (Day 3)**
**TN** recalled that a home care nurse came along and had a chat with the bear.

| | |
|---|---|
| TN: | Today's ... |
| Bear: | Oh, did you? |
| TN: | Today's nice weather, |
| Bear: | That's impressive. |
| TN: | On the veranda, the visiting ... |
| Bear: | Sounds great! |
| TN: | the visiting nurse... |
| Bear: | Sounds great! |
| TN: | you chatted with her, didn't you? |
| Bear: | Sounds great! |
| TN: | You had a good time, didn't you? |
| Bear: | I knew it. |

**Responding speech patterns**
The types of response speech were words with eight different functions (*Table 1*), as proposed by Kita & Ide (2007). All nineteen messages were recorded by a male native speaker of Hiroshima dialects who was around 60 years old. The local dialect was used to encourage the elderly person to continue the dialog (Aono et al., 2007). The recorded audio files were associated with prosody groups and pitches and stored in the memory of the HEARTalk board.

**Trial with an older woman**
The research participant was a woman in her 70s (TN) living alone in Hiroshima Prefecture. She interacted actively outside the home until she lost her partner three months before the experiment. Since then, she spent most of her day quietly at home. She had many friends and no cognitive issues. She had currently few opportunities to speak and was worried about the deterioration of her speech and swallowing functions. The researchers asked her to have conversations with the talking teddy bear for as long as possible, even though it wasn't clear whether the bear would react as expected.

The bear was placed on the dining room table where TN spent most of her time at home. She spoke to the bear for one to four minutes evening for three days. The first recording began shortly after the researcher's explanation. She ended the conversation herself when she was satisfied with the conversation or when she felt sleepy. On the fourth day her grandchildren came to stay with her. No recording took place. The duration of the conversation with the robot was 1min 40sec (Day 1), 4min (Day 2) and 4min (Day 3). The prosody and pitch of the words spoken by TN were inferred from the bear's responses. TN was asked to speak freely about her impressions of the bear, both just after the first trial, and on the first day of a week later when the researcher visited her.

**Analysis**
Speech analysis was performed by transcribing the conversations recorded during the first three days and categorising TNs' tone of speech and content. Two weeks after the first trial, the transcribed results were presented to TN on a spreadsheet. She was then asked to rate the appropriateness of each of the bears' responses on a 5-point scale from 1 to 5 (1 = not at all appropriate to 5 = very appropriate).

## Results

**Conversation**
TN spoke 33 (Day 1), 73 (Day 2) and 66 (Dat 3) sentences or phrases to the bear.

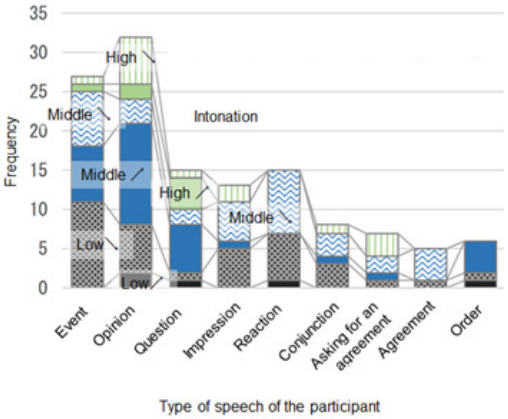


*Figure 2. Type of speech of TN. Colors indicate voice pitch (Black: Low, Blue: Middle, Green: High), patterns indicate the type of sentence ending (Solid: Up, Pattern: Down).*
*Event=events of the day;*
*Opinion=state her opinion or reaction to the bear's reaction;*
*Question=ask what the bear thinks;*
*Impression=convey impression;*
*Reaction=receptive response;*
*Conjunction=so, but, etc;*
*Asking for an agreement=..., right?, etc.;*
*Agreement=right., etc.;*
*Order=telling the bear to '~do~'.*

TN welcomed the teddy bear and nicknamed herself 'Kuma-san'. She told the researchers that she was delighted to hear the bear's voice in a familiar dialect and was intrigued by how the teddy bear's reaction managed the conversation in both predictable and unexpected ways. Subjects of the conversation were events of the day, her thoughts and opinions about the bear's answers, or questions for the bear. Some dialogue samples are shown in *Table 2*.

**Voice analysis**

TN used eight different voice types (*Figure 2*). TN often started the conversation by talking about the events of the day. There were more Opinion-related than Event-related utterances (*Figure 2*), indicating that the conversation was not one-sided, but progressed according to the bear's reaction. Questions often related to how the bear was doing, while trying to establish a conversation in a friendly tone.

For prosody, we found that Event often used descending prosody, while Opinion used both ascending and descending prosody at different pitches from low to high. This means that TN spoke in different tones by predicting bear's reaction.

TNs ratings of the bear's utterance after shown the transcriptions varied from 'good' to 'not understood'. The average rating was around 3 (fair). There were times when the bear responded to a puff of air in the middle of the conversation and the bear's utterance interrupted the conversation.

**Acceptance**

TN called the bear 'Mr. Bear', but its real name is 'Kumagoro'. She said that he speaks slowly in the Hiroshima dialect, which is friendly, that the eyebrows indicate that the bear probably speaks like a human male, and that its fluffy texture is pleasant.

TN said: *"I used a different part of my brain than I am used to because the bear's responses were often unexpected."* She noted that the bear was different from her pet dog, who could not speak but would do whatever she wanted. At first, she talked one-sidedly, but she soon realized that she needed to have a strategy to get the responses she was expecting. She managed the conversation by making up for the robot's lack of intelligence.

**Facilitating human communication**

TN enjoyed showing the bear to each of her visitors, including nurse, repairman and gardener. She let also visitors talk to the bear. This made it easier for her to make conversation with people she had never met before. Showing her visitors a novel object was common and usually lead to a conversation about the object. The conversation that is initiated between the robot and the visitor is a means of stimulating a conversation between the TN and the visitor about the conversation.

In addition, TN sent each recorded session to her daughters, who live apart from her. After the three-day session, the daughters reported that their mother's voice had become more expressive. After the experiment, TN continued to use the bear and eventually began to go out and participate in social activities again.

## Conclusion

We have successfully developed a talking teddy bear robot that brings joy to an elderly woman living alone, using only phonological information. The human-robot conversation was fully controlled by the user and encouraged sociable conversation. The limitations of this study are that the robot was used by only one person and for a short period of time. No specific improvements were made to the quality of the robot's responses. Future work will investigate whether deeper creative conversations are possible with a robot equipped with advanced Artificial Intelligence (AI).

**Research funding**
This research was funded by the Hiroshima International University Ordinary Research Fund.

**References**

Aono, A., Oki, T., Kagawa, N., Aoki, M., Miyake, M., Waki, T. (2015). The Role of Dialect and Movements in Robot-Communication. Journal of the Faculty of Human Cultures and Sciences of Fukuyama University 15:12-25 (in Japanese).

Cabinet Office, Government of Japan (2029). Aging Society White Paper, 2019 edition.

Nishio, M., Niimi, S. (2004). Relationship between Speech and Swallowing Disorders Following Neuromuscular Disease. Folia Phoniatrica et Logopaedica 56:291-304; https://doi.org/10.1159/000080066.

Inoue, E., Yamamura, S., Saito, Y. (2014). Effects of intervening vocalization on swallowing function in elderly individuals living in residential care facilities. Journal of the Ochanomizu Association for Academic Nursing 9(1):36-50.

Kita, S., Ide, S. (2007). Nodding, aizuchi, and final particles in Japanese conversation: How conversation reflects the ideology of communication and social relationships. Journal of Pragmatics 39:1242-1254; https://doi.org/10.1016/j.pragma.2007.02.009

Price, K. A., Tinker, A. M. (2014). Creativity in later life. Maturitas 78(4):281-286; https://doi.org/10.1016/j.maturitas.2014.05.025

Yamaha Corporation (2017). HEARTalk™ News Release, 2017.05.11 (in Japanese). Retrieved on 24 August 2023 from: https://www.yamaha.com/ja/news_release/2016/16051101/pdf/1605110101.pdf