International Journal of Web Information Systems

Implicit communication robots based on automatic scenario generation using web intelligence

Ryo Mashimo Tatsuya Kitamura Tomohiro Umetani Akiyo Nadamoto

Received 7 April 2016
Revised 9 June 2016
Accepted 10 June 2016

Implicit communication robots based on automatic scenario generation using web intelligence

Ryo Mashimo, Tatsuya Kitamura, Tomohiro Umetani and Akiyo Nadamoto
Konan University, Kobe, Japan

Abstract

Purpose – This paper aims to propose a system that generates dialogue scenarios automatically in real time from Web news articles. Then, the authors used the Manzai metaphor, a form of Japanese traditional humorous comedy, in the system. The generated Manzai scenario consists of snappy patter and a humorous misunderstanding of dialogue based on the gap of our structure of funny points. The authors create communication robots to amuse people with the generated humorous robot dialogue scenarios.

Design/methodology/approach – The authors propose the following: how to generate funny dialogue-based scenario from Web news and Web intelligence, automatically? How to create direction of robots based on the pre-experiment? The authors conducted experiments from three viewpoints, namely, effectiveness of Manzai scenarios as content, effectiveness of Manzai-Robots as a medium and familiarity of Manzai-Robots.

Findings – In this paper, the authors find two points, namely, the new communication style called "human–robots implicit communication-and bridging the knowledge gap using Web intelligence, to communicate smoothly between humans and robots.

Originality/value – Numerous studies have examined communication robots that mutually communicate with people. However, for several reasons, communicating smoothly with people is difficult for robots. One reason is the problem of communication style. Another is knowledge gaps separating humans and robots. The authors propose a new communication style to solve the problems and designate the communication style based on dialogue between robots as "human-robot implicit communication". The authors then create communication robots to communicate with people naturally, smoothly and with familiarity according to their dialogue.

Keywords Web media, Dialogue, Human-Robots interaction, Web intelligence

Paper type Research paper

1. Introduction

The rapid progress of robotics has supported the research and development (R&D) of robots of many kinds. Especially, R&D for communication robots is being pursued actively, producing robots that can converse with people. Many communication robots have been developed, such as AIBO[1], ASIMO[2] and Pepper[3]. Nevertheless, it is difficult for robots currently to communicate smoothly with humans. There are several

This work was partially supported by Research Institute of Konan University and JSPS KAKENHI Grant Number 26330347.

International Journal of Web Information Systems
Vol. 12 No. 3, 2016
pp. 312-335

1744-0084
DOI 10.1108/IJWIS-04-2016-0017

reasons for that difficulty, including problems of communication style and knowledge gaps separating human and robots. We particularly examine two points, new communication style and bridge the knowledge gap, for smooth communication between humans and robots.

1.1 New communication style

Kanda *et al.* (2002) studied human-robot communication. They described ways in which people feel friendly with respect to robots merely by watching their mutual dialogue. We designate the communication style based on dialogue between robots as "human-robot implicit communication". We then create communication robots to communicate with people naturally, smoothly and with familiarity according to their dialogue.

1.2 Bridge the knowledge gap

Many aspects of knowledge are related to the internet. Many studies have been conducted to extract knowledge from the internet. We consider knowledge related to the internet as useful for bridging a gap separating knowledge of humans and robots. We propose an automatically created robot dialogue scenario based on knowledge extracted from the internet.

We then create communication robots to amuse people through their dialogue. At this time, a usual robot dialogue is not attractive in terms of contents for users. One supposes that if the robot dialogue were humorous, it might then become more attractive to users.

Japan has a traditional type of humorous comedy called "Manzai": a kind of stand-up comedy. Manzai is typically a presentation by two comedians delivering a humorous dialogue routine. In Japan, Manzai has remained extremely popular over time. A Manzai show is now broadcast on TV every weekend. People continue to have strong familiarity with Manzai. The Manzai metaphor is, therefore, useful to create acceptable humorous robot dialogue. We propose the automatic generation of a Manzai scenario consisting of humorous dialogue generated automatically from Web contents.

In Japan, usual points of humor are snappy patter and misunderstandings. Our generated Manzai scenarios include dialogue of that type. Matsumoto (GansoBakushoOu, 2008), a professional Manzai scenario writer, said that "Manzai" scenarios based on news articles are the easiest for people to understand". Given that assumption, we generate Manzai scenarios based on news articles from the Web using knowledge of many kinds such as word ontology, extraction of sentiment and similar words obtained from searches of the internet.

The flow of the Manzai scenario generation system is the following:

- A user inputs a keyword to the system.
- The system selects a news article that includes key words from the internet.
- In real time, the system generates a Manzai scenario consisting of humorous dialogues based on knowledge from the Web.
- The robots "Ai-chan and Gonta" (Plate 1) perform a Manzai routine in real time.

Plate 1.
Manzai-Robots: ai-chan (left) and gonta (right)

When generating a Manzai scenario as friendly content for humans, it is important not only to generate funny dialogue but also to generate the directions, which are the voice, actions, and so on between robots.

As presented in this research, the technical points are the following:

- How to generate funny dialogue automatically in real time using Web intelligence?
- How to generate direction?

We already proposed how to generate funny dialogue automatically (Mashimo *et al.*, 2015; Mashimo *et al.*, 2016). In this paper, we mention about the generation of funny dialogue automatically and propose how to generate direction. Furthermore, we conducted experiments from three viewpoints:

(1) effectiveness of Manzai scenarios as content;

(2) effectiveness of Manzai-Robots as a medium; and

(3) familiarity of Manzai-Robots based on comparison between Manzai-Robots and other media.

The remainder of this paper is organized as explained below. First, Section 2 presents a summary of related works. Section 3 explains the basic concept. Section 4 presents a description of Manzai scenario generation. Section 5 presents directions related to robot's voice. Section 6 presents discussion related to experimental evaluation, in addition to evaluation of results. Finally, in Section 7, we end this report with conclusions and expectations for future work.

2. Related work

Hayashi *et al.* (2005) proposed a "robot Manzai" system for human-robot communication. This system realized "Manzai" using a pair of robots. A comparison experiment was conducted between "robot Manzai" and "Manzai" shown in a video performed by humans. Consequently, Hayashi *et al.* demonstrated the usefulness of "robot Manzai" as entertainment. Their robots specifically examine the human action in "Manzai", but we particularly examine the scenario as Manzai contents.

Our proposed system is important also in that it facilitates the understanding of news contents. Park *et al.* (2009) developed the "News Cube", a news browsing service that mitigates media bias. Kitayama and Sumiya (2007) proposed a new search method of retrieving comparison contents for news archives. Our proposed system is aimed at an understanding of news contents by the generation of humorous dialogue based on news articles.

Numerous studies have presented and assessed dialogue analysis. Ishizaki and Den (2001) present a good summary of work that has been reported in this area. Their approaches include analyses of real-world dialogues and extraction of intentions from the dialogue. Our efforts include the generation of humorous dialogue from internet news articles. Numerous studies have examined conversational agents, such as those reported by Bouchet and Sansonnet (2009) and Ishii *et al.* (2013). In almost all such studies, CG characters communicate with humans based on dialogue. Applications of these studies are used in education, entertainment, communication support and other fields for communication with CG characters by humans. Because of the main purpose of our research, however, people are expected to be entertained and healed by watching a robot dialogue based on our proposed automatically generated scenario.

Our proposed automatic generation of Manzai scenario is a kind of computational humor. Yamane and Hagiwara (2010) propose generation of funny proverbs based on the impression and length of a word using n-grams. Takizawa *et al.* (1996) propose the automatic generation of Dajare, a Japanese sort of pun. Our proposed method, the automatic generation of humorous dialogue from Web news, differs from these efforts. Especially, one characteristic of our proposed method is the generation of humorous dialogue based on gaps of two types: sentiment gap and rival gap, which differ from other work related to computational humor.

Research undertaken in the field of robotics has investigated many entertainment robots, as exemplified by studies reported by Shiomi *et al.* (2008); Khosla and Chu (2013); Rae *et al.* (2013); and Yamaoka *et al.* (2006).Shibata (2010) proposed Paro, which is an entertainment robot and it heals people with his cute gestures. PALRO, ifbot, KIROBO and others communicate with people using cute dialogue. Tanaka *et al.*' (2006) study about entertainment robots for children is based on gesture and dance of a small bipedal robot. They talk with people, but our Manzai-Robots mutually converse. People need only to watch their humorous dialogue to be entertained.

3. Basic concept

3.1 Structure of humorous points

Abe (2006) has described the structure of funny points that consist of gap separating concepts (show Figure 1). He says that the concept gap is an important aspect of the humorous points. We specifically examine that point. Then, we describe the structure of our funny points for dialogue based on Abe's structure. In our structure, the original

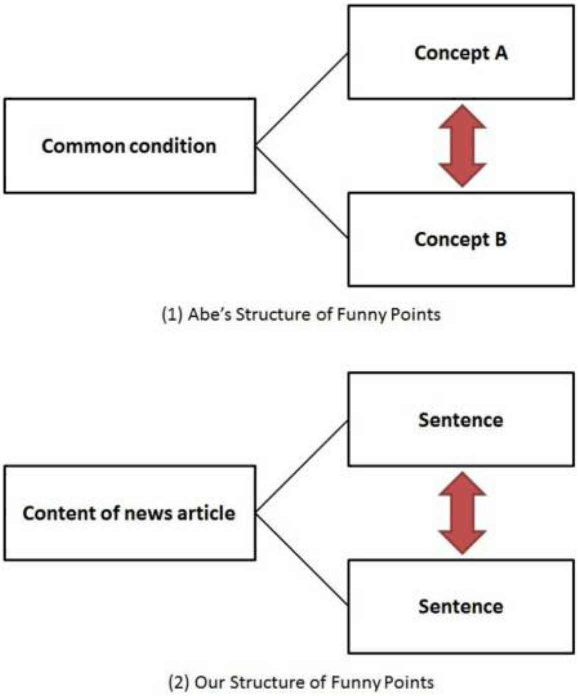

Figure 1.
Structure of funny point

sentences of Web news articles are Abe's common condition. A word is a concept in the original sentence A. A word that is extracted by our system is concept B. We propose gaps of two types, which are important for funny dialogue. One is a rival relation; the other is a sentiment gap. In the case of a rival relation, we extract rival words from the Web such as "baseball" and "soccer". For a sentiment gap, we extract a sentiment from the original sentence. Then, we create a new sentence having the opposite sentiment. As described herein, we propose a means of extracting rival keywords from the Web to create a sentiment gap from the Web. Then, we use our proposed technique to generate a Manzai scenario for presentation by robots.

3.2 What is Manzai?

As described in this paper, we propose an automatic Manzai scenario generation system that produces humorous dialogue from Web news articles. Japan has a traditional form of comedy routine called Manzai, which typically consists of two comedians performing humorous dialogues. It is similar to "stand-up comedy" in English or "xiang sheng" in Chinese. It usually includes two performers: one is the "boke", or stooge; the other is the "tsukkomi", or straight man. The boke says things that are stupid, silly or out of line, sometimes using puns, whereas the tsukkomi delivers quick, witty and often harsh responses. Regarding our Manzai-Robots, Ai-chan (left side of Plate 1) is the tsukkomi (straight woman), and Gonta (right side of Plate 1) is the boke (stooge). Furthermore, the Manzai scenario has a three-part structure: the introduction, the body and the conclusion.

3.2.1 Introduction part. The introduction part consists of several short dialogues. In our system, it consists of a greeting and a discussion in which the theme of the original Web news is presented first.

3.2.2 Body part. The body part, the main part of the Manzai scenario, consists of humorous dialogue. We designate the body part as the block of humorous dialogue called the dialogue component, which is created automatically based on a sentence in a Web news article. Consequently, we create a dialogue component that is a set of humorous dialogue from a sentence in a Web news article. We generate humorous dialogue based on the structure of funny points of dialogue using gaps of two types, which are rival relations and sentiment gaps (Figure 1).

3.2.3 Conclusion part. The conclusion part includes a farewell and a final laughing point. In our system, the conclusion part is a riddle of words related to Web news. Figure 2 portrays an image of correspondence of Web news article and automatically created scenario.

3.3 Manzai-Robots

Plate 1 portrays our Manzai-Robots: "Ai-chan and Gonta". Ai-chan is 100-cm tall. Gonta is 50-cm tall. They have laptop PCs inside them. Ai-chan is a server. Gonta is a client. Ai-chan, connected to the internet, generates Manzai scenarios in real time. After the generation of Manzai scenarios, they perform a Manzai routine based on the scenario. They mutually communicate using wireless communication. Ai-chan manages their speech control. Their eyes are small displays. They can change their eye appearance based on their scenario. They manage speech generation, robot motion and eye expressions independently.

4. Generation of Manzai scenario

When a user inputs keywords for Manzai to be watched, we generate a Web news-based Manzai scenario automatically based on the three parts of actual Manzai routines, namely, the introduction part, body part and conclusion part. In each part, we generate humorous dialogue from Web news articles with related information obtained from the internet. We transform declarative sentences of Web news articles into humorous dialogue using related information from the internet based on our proposed gaps of two types.

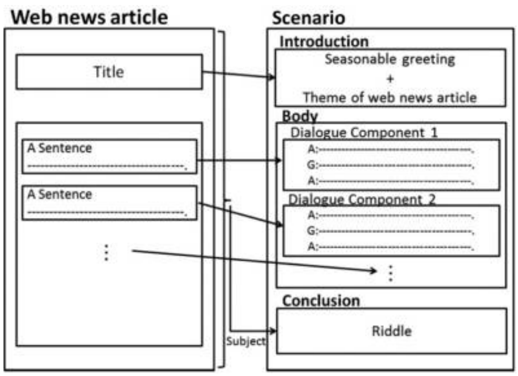

Figure 2. Image of correspondence of Web news and scenario

4.1 Introduction part

The introduction part consists of a first laughing point with a greeting and a connection to the theme of Manzai to the main part. We create a season dialogue database in the system and choose dialogue from the database. In December, for example, the introduction part of the Manzai has a greeting for December, and Web news title is "Nobel award ceremony!!" as shown below, where A denotes Ai-chan and G stands for Gonta:

G: Hello Everyone. It's December. Christmas season!

A: Ah yes, Christmas season is coming. Anyway, have you heard the news?

G: Hmm[…] […] "Nobel Prize awards ceremony!!" I did not know that.

A: That sounds interesting. Let's start a Manzai routine about the Nobel Prize awards ceremony!!

4.2 Body part

Our automatically created scenario consists of multiple components in the body part.

One component is one funny technique. We propose funny techniques of four types, which are "exaggeration", "word-mistake", "rival-mistake" and "sentiment-mistake". Our structures of funny points are the gaps separating dialogue. We generate humorous dialogue based on our structure of funny points of dialogue (Figure 1) using gaps of four types: number gaps, topic gaps, rival relations and sentiment gaps.

4.2.1 Exaggeration component. The first step in generating dialogue using exaggeration is to use impossible (larger or smaller) numbers. This component is a number gap. This humorous technique is sometimes used ordinarily in dialogue routines in Japan. We feel familiar about such types of exaggeration. When a sentence includes numbers, the system increases the numbers by some substantial and unbelievable factor. For example, if a sentence is "the audience of the games is 50,000 people.", then the exaggerated dialogue might become the following:

G: In this year, 300 Japanese got Nobel Prizes!!

A: Wahoo. 300 Japanese? That is way too many people!!

G: Oh, I made a mistake. Three Japanese got Nobel Prizes this year.

4.2.2 Word-mistake component. This component is an intentional word-mistake in the dialogue. This is topical (word) gap. A typical word-related mistake is to change a single word to another word based on changing only one character in a word such as "nose" and "hose". For example:

G: Do you know blue TEDs were crucially important to create a full color image?

A: Blue TED? TED is a global set of conferences run by the private non-profit Sapling Foundation, under the slogan "Ideas Worth Spreading". LED is right.

G: Oh, I made a mistake. It is blue LEDs.

The Manzai-Robot has only one dictionary, a Japanese dictionary for children in the server (Ai-chan) machine. We use the Japanese dictionary to extract mistaken words such as "TED". Japanese consists of vowels and consonants. It is a simple matter to change the first consonant to another one. If there is a word for which the first consonant in the dictionary can be changed, it becomes the candidate of the mistaken word. For example, In Japanese, "Touhyou (vote)" is change to "Kouhyou (favorable comment)".

After a word-mistake, Ai-chan, the tsukommi, emphasizes on the mistaken words by explaining the mistaken word. In the example given above, Ai-chan explains the TED as "TED is a global set of conferences run by the private non-profit Sapling Foundation, under the slogan 'Ideas Worth Spreading'". This explanatory sentence comes from the first sentence of a Wikipedia entry.

The reason, as Nakayama (2007) has reported, is that the first sentence of Wikipedia is usually a presentation of the word definition. Furthermore, the sentence is usually a formally worded sentence. For that reason, it is funny to create a gap separating such formal language and the otherwise casual humorous dialogue. Next, Gonta, the boke, realizes its own mistake and sets it right.

4.2.3 Rival words gap based dialogue. We infer that if the words which are in dialogue have a mutual rivalry, then we change the original word in a sentence to a rival word, whereby the dialogue becomes a sentence that includes a misunderstanding. Then, it becomes a funny point. Our proposed rival words are "Tokyo" and "London", and "baseball" and "soccer". The two words are contrasting pairs. Then, we extract a word that is a rival to the keyword and change the word from the keyword to the rival word. The following presents an example of the dialogue based on a rival mistake:

G: The Olympic Committee chose the 2020 Olympic venue as London.

A: Is that right? London? You have made a mistake of some kind. London was the last Olympic venue. You have made a mistake, haven't you?

G: Oh no! I have made a mistake. Not London, it is Tokyo!!

Definitions of proposed rival words are the following:

4.2.3.1. They have the same upper ontology. The upper ontology of "Tokyo" and "London" is national capitals. The upper ontology of "baseball" and "soccer" is ball sports. Each pair shares the same upper ontology. When we extract the upper ontology, we use the hierarchy structure of Wikipedia. At this time, the upper ontology of a word is not just one ontology; usually, a word has multiple upper ontologies. We use all upper ontologies to extract rival words. The rival words are child words of the upper ontology. Then, many words can become candidate rival words. We regard the candidates of rival words as ranked. The top-ranking word becomes a rival word. At this time, we consider an upper ontology that has few child words as more important than an upper ontology that has many child words because the former has a higher instance level than the latter. We then calculate the ranking of a candidate of rival words using the following expression:

$$
\begin{aligned}
Sta(s_i) &= \log \frac{n}{N} \\
Rel(e_i) &= \sum_{i=0}^{m} Sta(s_i)
\end{aligned}
\tag{1}
$$

The given s_i denotes an upper ontology; $Sta(s_i)$ signifies weight of s_i. In addition, n represents a number of s_i's lower ontology; N is the number of words in the corpus; the given e_i stands for a candidate of rival word; $Rel(e_i)$ signifies the ranking weight. In addition, m represents a number of e_i's upper ontology, which is the same as the keyword. As described herein, we use $N = 2{,}931{,}465$ words.

4.2.3.2. They have a similar degree of recognition. The candidate rival words of "baseball" are "soccer" and "futsal". At this time, the degree of recognition of soccer is more similar to that of baseball. Therefore, we regard soccer as better than futsal for use as a rival word of baseball. That is, the two words have similar degrees of recognition.

We regard the degree of recognition as the number of results obtained from a Web search. We calculate the similarity of the degree of recognition $Con(\text{key}, e_i)$ between keywords *key* and e_i as follows:

$$Con(key,\ e_i) = 1 - \frac{log\, Cog(key) - Cog(e_i)}{max\{Cog(key),\ Cog(e_i)\}} \tag{2}$$

In that equation, $Con(key)$ is the number of results of a Web search using word *key*; $Con(e_i)$ is the number of results of a Web search using word e_i. After calculating $Rel(e_i)$ and $Con(e_i)$, we regard the result of geometric mean between $Rel(e_i)$ and $Con(e_i)$ as a ranking weight. The word having the highest ranking weight becomes a rival word. Subsequently, we create the rival-mistake component based on changing the keyword to the rival word and generate a misunderstanding humorous dialogue. Table I presents an example of the rival words we extract.

4.2.4 Sentiment gap-based dialogue. We propose sentiment mistake gaps of two types: a word sentiment mistake type and a sentence sentiment mistake type.

4.2.4.1 Word sentiment mistake. We infer that when the sentiment of the word is mistaken, it is interesting. At this time, we use the Japanese traditional technique called "nori-tsukkomi". In a nori-tsukkomi, first, the boke (stooge) states some outrageous sentence and plays the clown. Next, the tsukkomi (straight man) gets into line with the boke's non-sensical sentence. Then, the tsukkomi sets the sentence right and makes a fool of both the boke and itself. Under such circumstances, three techniques are used for dialogue generation, as described below:

Q1. How to create the first outrageous statement?

Q2. How does the tsukkomi agree with the boke's line?

Table I.
Rival-words are related to keywords

Keywords	Rival-words
Italy	France
Japan	China
Baseball	Soccer
Chess	Reversi
Dog	Rabbit
Barack Obama	George Washington
Internet	Newspaper

Q3. How does one create a corrective sentence?

We use the word-mistaken technique in equation (1). The word-mistaken technique is where the single character is changed, producing a different word, such as "hose" and "nose". In equation (2), we use a sentiment of the mistaken word in equation (1). At this time, we extract sentiment words that co-occur with mistaken words from the Web and calculate the co-occurrence ratio. We specifically search the Web pages using the word that we want to extract a sentiment as a keyword. Then, we extract sentiment words, which are adjectives, from the snippet of the results. Then, we calculate the co-occurrence ratio between the word and each sentiment word. The sentiment word having the highest co-occurrence ratio becomes the sentiment of the mistaken word. In equation (3), we use the original sentence in Web news article. An example of nori-tsukkomi follows:

G: He waters the garden with his nose.

A: That's right!! His nose is long. Humm[…]
No way!! A nose means "Anatomically, a nose is a protuberance in vertebrates that houses the nostrils, or nares, which admit and expel air for respiration in conjunction with the mouth. You have made a mistake! It is not a nose, it is his hose!!"

G: Oh!

In equation (2), there is agreement based on the fact that his nose is long and his impression word of nose is "long", but after that agreement, Ai-chan mentions the mistaken word. At this time, we extract a sentence mentioning the mistaken word from Wikipedia. The Wikipedia sentence is usually a hard and formal sentence that emphasizes a humorous contrast with the humorous dialogue.

4.2.4.2 Sentence sentiment mistake. We consider that the sentence in a dialogue has opposite sentiment to the original sentence. It is a misunderstanding of dialogue. It becomes a gap of structure of funny points. For example:

G: Three Japanese were awarded the Nobel Prize in Physics for the invention of blue light-emitting diodes (LEDs).
It is very sad news.

A: What? Why do you think the news is sad?

G: Because three Japanese awarded the Nobel Prize in Physics. I feel very sad.

A: Do you want to be awarded the Nobel Prize in Physics? Are you so smart?

We calculate the sentiment of the sentence and change the word which has opposite sentiment of the original sentence. Then, Ai-chan corrects the sentiment of the Gonta. In the case above, the correct first sentence of the Gonta is "Three Japanese awarded the Nobel Prize in Physics for the invention of blue light-emitting diodes (LEDs)". It is good news in a Web news article. We change the "awarded" to "lost" and thereby radically alter the sentiment of the sentence. When we generate dialogue based on sentiment mistake gap, we use multidimensional sentiment, which includes three bipolar scales, as proposed by Kumamoto (2010). Kumamoto's three bipolar scales are sufficient to calculate Web news article sentiments because he creates the bipolar scales using news articles. The bipolar scales are "Happy-Sad", "Glad-Angry" and "Peaceful-Strained". Generating a sentiment mistake gap-based dialogue is done as follows:

- We calculate the sentiment of each sentence in a Web news article using Kumamoto's sentiment extraction tool. At this time, the results are sentiment values in each bipolar axis. We normalize the results, such as a value of 1.0-0.0 means left side axis and a value of 0.0-1.0 denotes the right side axis. For example, when the result is "happy-sad" as 0.12, "glad-angry" as 0.26 and "peaceful-strained" as −0.07, the sentiment value of the sentence becomes happy as 0.12, glad as 0.26 and strained as 0.07.
- We infer that the highest sentiment value of a sentence is a sentiment A_e of the sentence. In a case of equation (1), the glad becomes the sentiment of the sentence.
- We extract the word W_e which has highest value of A_e. Then, we extract antonym T_e of W_e from the antonym corpus.
- We create a new sentence using T_e.

4.3 Conclusion part

The conclusion part consists of a farewell and the final laughing point. We use an automatically generated riddle in our conclusion part. Japanese people are familiar with riddles. They have been used for entertainment for a long time. In Japanese, riddles are usually "How are X and Y similar?" The answer is typically some form of homonym or pun: Z and Z'. For example:

G: What is the similarity between soccer and baby?

A: I do not know. What?

G: Both of them have a ball ("bawl")!!

In this example, soccer is X, baby is Y, ball is Z and bawl is Z'. We first extracted the word X (soccer) from the Web news article. Next, we extracted words that have a high co-occurrence ratio to X from the internet as candidates for word Z. We next extracted homonyms of candidates of Z from the dictionary, which is inside AI-chan. If the candidate of Z has a homonym, then it becomes Z. The homonym becomes Z'. Subsequently, we extracted the word having the highest co-occurrence ratio to Z'. from the internet. It becomes Y. Then, we generated dialogue based on X,Y,Z and Z'. Figure 3 shows an automatically generated scenario from web news article.

5. Direction

Impressions of robots differ by the direction of the robots. Therefore, direction for the robots is an important to the Manzai scenario. The direction consists of action, voice and other characteristics that affect impressions. As described in this paper, as a first step of creating direction, we target voice direction.

5.1 Sentiment by voice

Moriyama (2009) has studied synthesis of speech, which can effect sentiment according to its pitch, speed and volume. We assume that pitch, speed and

Figure 3.
Example of generated scenario

volume to Manzai-Robots can be a sentiment that matches the flow of scenario. However, we consider that the variation of volume hinders understanding of the contents of the news. Then, we use pitch, range and speed to sentiment expression by voice.

5.2 Pre-experiment 1

We determined the parameters of pitch, range and speed for effectiveness of sentiment that Kumamoto's three bipolar scales ("happy-sad", "glad-angry" and "peaceful-strained"). We generated 20 types of parameters for which pitch, range and speed are in each of 75, 100 and 125 per cent from default, respectively. Table II presents the 20 types of parameters. The read out sentences are two types which are casual sentences and formal sentences. All are non-sentiment sentences. The sentences are shown in Table III. We listened to a total of eight participants: four participants listen to casual sentence and four participants listen to the formal sentence in each parameter. After participants listen to them, they judged six sentiments. When a participant felt some sentiment(s), they assigned one point to each sentiment.

5.2.1 Results and discussion. Table IV shows the results. In the case of happy, glad and peaceful, the same parameters are highest scores irrespective of the type of sentence. However, in the case of negative sentiments of sad, angry and strained, the results differed by the type of sentence. In the case of sad, No. 3 (pitch = 100 per cent, range = 75 per cent and speed = 75 per cent), No. 12 (pitch = 125 per cent, range = 75 per cent and speed = 75 per cent) and No. 18 (pitch = 75 per cent, range = 100 per cent and speed = 100 per cent) become high scores. However, we could not find them in common. In the case of Angry, No. 7 (pitch = 75 per cent, range = 100 per cent and speed = 125 per cent), No. 8 (pitch = 75 per cent, range = 125 per cent and speed = 125 per cent) and No. 16 (pitch = 75 per cent, range = 75 per cent and speed = 125 per cent) become high scores. In this result, the range is 75 per cent, and speed has become 100 or 125 per cent. In the case of strained, No. 2 (pitch = 125 per cent, range = 125 per cent and speed = 100 per cent), No. 4 (pitch = 100 per cent, range = 100 per cent and speed = 125 per cent), No. 6 (pitch = 125 per cent, range = 75 per cent and speed = 125 per cent) and No. 11 (pitch = 100 per cent, range = 125 per cent and speed = 125 per cent) become high scores. In this result, speed is 125 per cent, and pitch has become 100 or 125 per cent.

We were unable to find the common point of the parameter of different dialogue styles in all sentiments. We will use the six parameters which scored the most points, which is the sum of the scores for casual speech and polite speech in Table IV, when expressing the six sentiments when delivering news. When it comes to grief, Nos 3, 12 and 18 have each scored three points. Therefore, we will use No. 12, which had the greatest difference by default. We will use "shock" and "irritation" that was freely stated for the retorts of three types. Then, we will use "slowly" for asking riddles as parameters. Then, No. 10, which scored the highest points as "calm", will be treated in the same way. The ultimately chosen parameters are presented in Table V.

5.3 Pre-experiment 2

In this pre-experiment 2, we measure the following two points:

(1) The parameters selected by Pre-experiment 1 are appropriate.

(2) Participants feel active participant in the dialogue or not.

In this Pre-experiment 2, there are five participants. They first listened to the Manzai generated automatically. Their voice parameters are determined in Pre-experiment 1. Subsequently, they judged according to a five-point scale based on three viewpoints which are familiarity, comprehensibility and interest. Table VI shows

No.	Pitch	Range	Speed
1	100	100	100
2	125	125	100
3	100	75	75
4	100	100	125
5	125	75	100
6	125	75	125
7	75	100	125
8	75	125	125
9	125	125	75
10	100	125	75
11	100	125	125
12	125	75	75
13	75	125	100
14	125	100	100
15	100	125	100
16	75	75	125
17	100	75	100
18	75	100	100
19	100	100	75
20	75	75	75

Table II. 20 parameters used in the experiment

Casual speech		Polite speech	
Tsukkomi	When is Professor Kawada's class exam?	Boke	Are these meat balls ready to eat?
Boke	The Professor said no midterm	Tsukkomi	No, they are not ready to eat yet
Tsukkomi	Really?	Boke	The inside is not done yet
Boke	Sure. But, two finals	Tsukkomi	Yes, they would be better heated a little more

Table III. Examples of dialogue in non-sentiment voice

the Manzai scenario news titles and their sentiments. We compared the Manzai using sentiment voice and non-sentiment voice and asked participants if their impression changed. They responded on a five-point scale. Next, we asked participants the same question in the three kinds of tsukkomi part and riddle part. Furthermore, we asked the participants about the effectiveness of using a sentiment voice.

5.3.1 Results and discussion. Figure 4 presents results for each type of sentiment news based on three points: familiarity, comprehensibility and interest. In all cases, familiarity of using the sentiment voice becomes better than using a non-sentiment voice. The results of comprehensibility when using a sentiment voice are also better than using non-sentiment voice. Especially, the comprehensibility of strained is 100 per cent participants feel using the sentiment voice is better than using a non-sentiment voice. The results of interest are not better than other viewpoints. Tsukkomi and riddle in Figure 4 portray the results of comparing sentiment voice

Table IV. Casual speech and polite speech

	Happy		Sad		Glad		Angry		Peaceful		Strained	
	C	P	C	P	C	P	C	P	C	P	C	P
1	–	–	–	–	–	–	–	–	–	–	–	–
2	1.5	0	0	0	1.5	1.5	0	0	0	0	2.0	1.0
3	0	0	3.0	0	0	0	0	0	1.0	3.0	0	0
4	0	0	0	0	0	0	0	0	0	0	3.0	2.0
5	0.5	0	0	0	0.5	1.0	0	0	0	1.0	1.0	0
6	0	0	0	0	0	0	0.5	0	0	0	1.5	2.0
7	0	0	0	0	0	0	2.0	3.0	0	0	0	0
8	0	0	0	0	0	0	1.50	2.0	0	0	0.5	2.0
9	0.5	0	0	0	0	0	0	0	1.5	3.0	0	0
10	0.5	0	0	0	0	0.5	0	0	3.5	3.5	0	0
11	0	0	0	0	0.5	0	0.5	0	0	0	1.5	2.0
12	0	0	2.5	0.5	0	0	0	0	1.5	3.5	0	1.0
13	0	0	0.5	0	0.5	0	2.5	1.0	0.5	1.0	0	1.0
14	1.0	1.5	1.5	0	1.0	0.5	0	0	0.5	0	0	0
15	0	0	0	0	1.0	0	0	1.0	1.0	0	1.0	0
16	0	0	0	0	0	0	3.0	1.5	0	0	1.0	1.5
17	0	0	1.0	0	0	0	0.5	0	0.5	3.0	0	0
18	0	0	2.0	1.0	0	0	0	0	0	2.0	1.0	0
19	0	0	1.0	0	0.5	0.5	0	0	2.5	3.5	0	0
20	0	0	1.0	0	0	0	0	0	0	1.0	0	0

Notes: C = casual speech; P = polite speech

Table V. The parameters of voice of robots

Performer	Sentiment	Pitch	Range	Speed	No.
Boke	Happy	125	100	100	14
	Sad	125	75	75	12
	Glad	125	125	100	2
	Angry	75	100	125	7
	Peaceful	100	125	75	10
	Strained	100	100	125	4
Tsukkomi	Shock	75	125	100	13
	Little Irritation	75	125	125	8
	Irritation	100	125	125	11

and non-sentiment voice in three kinds of tsukkomi part and riddle part. In tsukkomi and riddle, the rates of “changed well” and “changed” were found to be high.

Figure 5 presents the results of effectiveness of using the sentiment voice. Almost half of the participants feel good about the effectiveness of sentiment voice, especially the case in which the riddle is most effective. However, the strained case shows that almost all participants felt bad. For the cases of anger and in tension, the participants were split on the answer. For example, some participants reported that if anger is revealed too much, then the person felt fear. The reason is that the parameter is not good. Therefore, we must consider additional details of the parameters.

News	Sentiment
Tanabe city, Wakayama and Kansai University Evaluate revitalization of depopulated areas	Happy
Private house on fire in Gunma Prefecture	Sad
Coming-of-age ceremony held in Amagasaki	Glad
Detention operation against Mexican drug lord	Angry
Tsukuba University students continues to send acorns picked up on campus to children in Fukushima	Peaceful
Partial preservation work on condominium at Fukuchiyama derailment site begins: ceremony for safety held	Strained

Table VI. Sentiment of news

6. Experiment

We evaluated the effectiveness of our proposed human–robot implicit communication through user experiments. Our proposed Manzai-Robots consist of "robots" and a "scenario". We measured three types of aspects:

(1) the effectiveness of a Manzai scenario as a "content for robots";

(2) the effectiveness of Manzai-Robots as "media"; and

(3) the effectiveness of types of familiarity for Manzai-Robots.

We conducted all experiments with each of 18 participants, all of whom were Japanese people in their 20s.

6.1 Experiment assessing the effectiveness of Manzai scenario content

6.1.1 Condition of the experiment. In our experiment, we measured the effectiveness of the Manzai scenario using different contents and other conditions: robots (visual) and synthesis of speech (sound).

6.1.1.1 Manzai-Robots. The robots have two actions, facing front and turning the robot face to another robot, which is the speech scenario. The robots have 50 eye varieties. They change eye expressions based on the scenario sentiment. The time of a Manzai exchange is about 3 min.

6.1.1.2 News-robots. First, the robot (Ai-chan) reads out the news title. Subsequently, another robot (Gonta) reads out the entire main news article. The robot speech synthesis is the same for Manzai-Robots. The robots do not move when they read a news article because ordinarily news anchors do not move when they read news. The reading time is about 30 s or 1 min.

First, participants watched Manzai-Robots and answered the questionnaires. Thereafter, they watched the News-robots and also answered questionnaires. The questionnaires, which are generated by viewpoints of understandable, interesting and familiarity, are the following:

Q1. Are the Manzai-Robots/News-robots understandable?

Q2. Are the Manzai-Robots/News-robots interesting?

Q3. Are the Manzai-Robots/News-robots familiar?

Q4. Which would you prefer watching to get today's news: Manzai-Robots or News-robots?

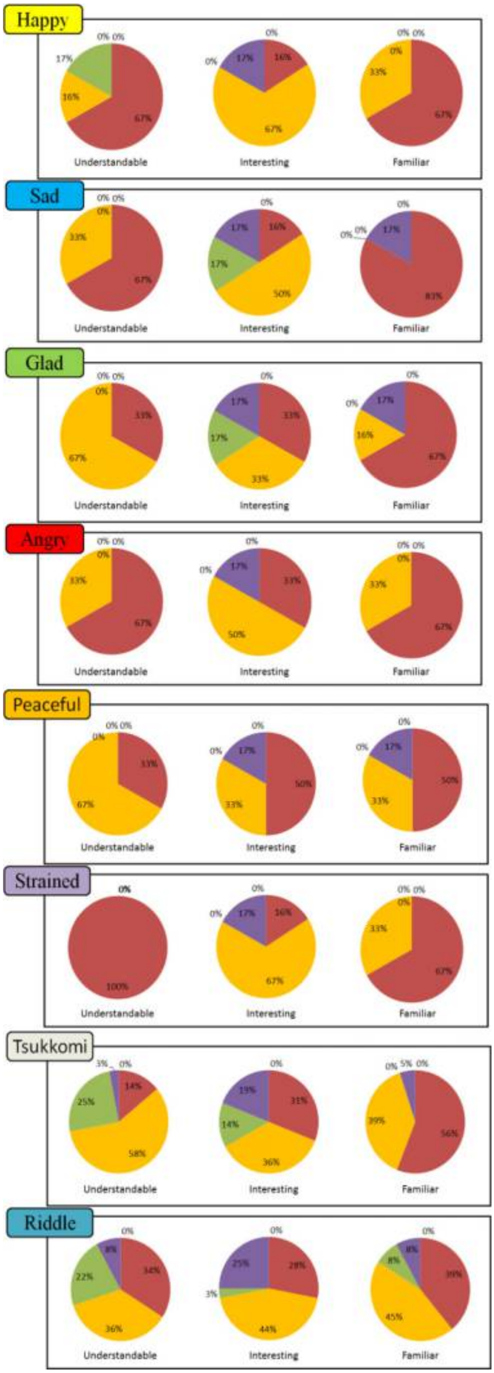

Figure 4.
Results for each type of sentiment news based on three points

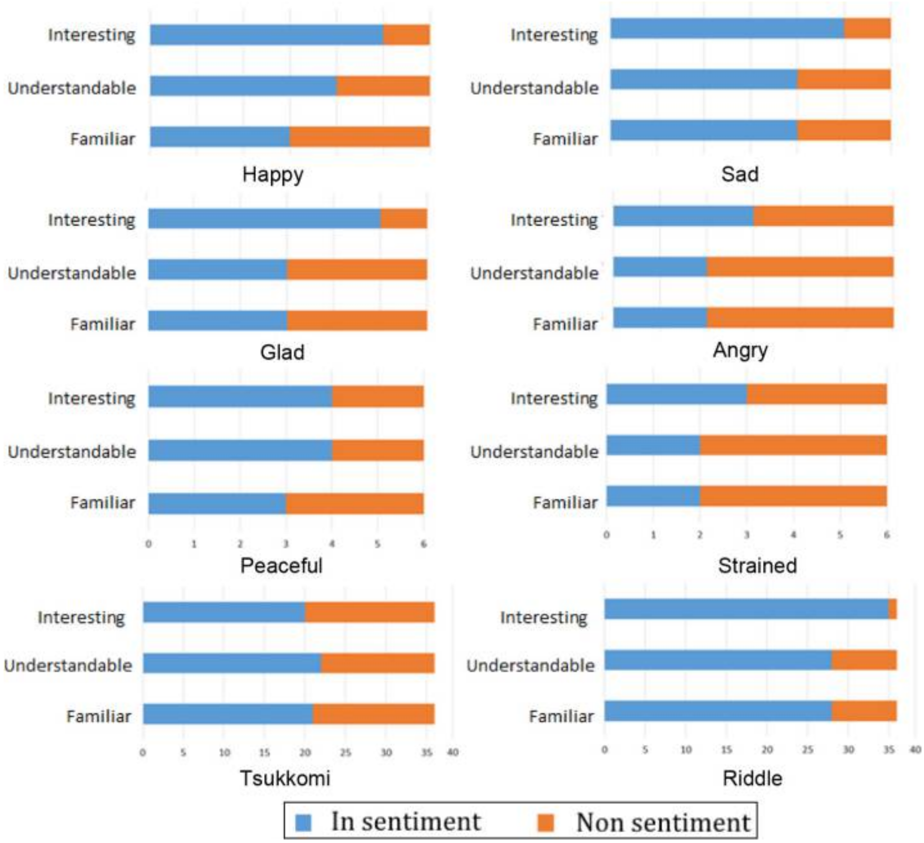

Figure 5.
Results of effectiveness of using the sentiment voice

Responses to the questionnaires were given using a five-point scale (5, highest; 3, middle; 1, lowest). The last questionnaire presents a three-fold choice (Manzai-Robots, Reading robots, neither).

6.1.2 Results and discussion. Experimental results are presented in Figure 6(a). Manzai-Robots are more understandable, more interesting and more familiar than News-robots. The body of Manzai-Robots and News-robots are the same, but their interaction type and content differ, which means that the Manzai style interaction robots are more effective than news-style interaction robots. The five participants reported feeling that they have difficulty in concentrating with News-robots; three participants feel some rigidity in the manner of the News-robots. The results of *Q4* show that 16 participants liked Manzai-Robots, one participant liked News-robots and one participant liked both. A participant responded that a News-robot was more effective than a Manzai-Robot because the time to watch Manzai-Robots was longer than that of the News-robots. In the near future, we must consider the Manzai playing time. We discuss the familiarity of Manzai-Robots from viewpoints of two types, namely, the dialogue of Manzai and the direction of robots.

6.1.2.1 Viewpoint of dialogue of Manzai. Participants reported that they can understand news information from Manzai-Robots, and that the dialogue between robots is more familiar and interesting than merely reading out the news. Furthermore, seven participants reported that the dialogue of Manzai-Robots sometimes includes a "boke (stooge)", and that it is more familiar. Participants reported that they can watch Manzai-Robots with more concentration than the News-robots because they feel that the content is interesting. Results show that funny dialogue makes people feel more familiar.

6.1.2.2 Viewpoint of direction of robots. The perspective of direction signifies robot behavior and expression. Both Manzai-Robots and News-robots are passive media, but participants feel more familiarity to the Manzai-Robots. They said the rationale is that the Manzai robot not only says funny dialogue but also moves its face frontward to the participants and turns the robot face toward another robot for the dialogue scenario. Moreover, the eyes change so that they look cute. Therefore, the robot production is also an important parameter that deserves some attention to produce attractive robots.

We conducted paired-sample *t*-test to assess comprehensibility, interest and familiarity in Manzai-Robots and News-robots as a statistical test. Table VII presents results of *t*-tests. A significant difference was found in the scores for understandable $t(25) = 2.208$, $p = 0.03$, for interesting $t(27) = 5.827$, $p = 0.0$, and for familiar $t(29) = 5.395$, $p = 0.0$. One finds a significant difference between Manzai-Robots and News-robots in all scales, which are understandable, interesting and familiar. The average value of Manzai-Robots is higher than that of News-robots. These results suggest that Manzai-Robots are more effective than News-robots in all scales. Especially, before we conducted the experiment, we consider the News-robots as more understandable than Manzai-Robots, because Manzai-Robots (scenario) append more information such as boke and tsukkomi to original news and people lose real news articles. However, the results suggest that Manzai-Robots are more understandable than News-robots. Results demonstrate that people have a more engaging experience when hearing funny dialogue that highlights the important words in the scenario. Results show that Manzai-Robots (scenario) provide an effective means to deliver content. Results show that the Manzai scenario for robots is effective for making people feel familiarity with the robot.

6.2 Experiment of effectiveness of Manzai-Robots as media

We also measured the effectiveness of Manzai-Robots as media using the same scenario and different media which are robots and other media. In our manuscript, we use CG animation as another medium.

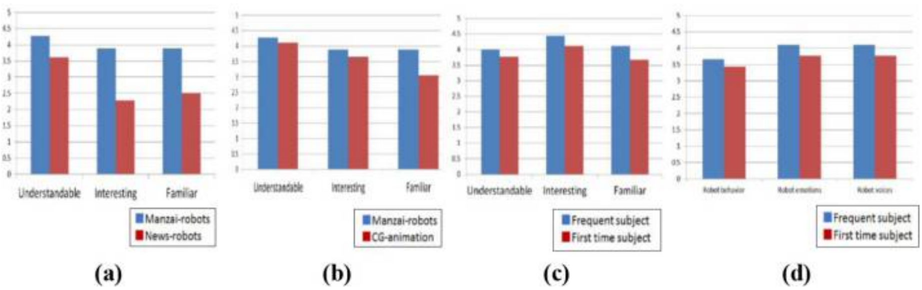

Figure 6.
Result of comparison between Manzai-Robots and News-robots

6.2.1 Conditions of the experiment. We used the TV program Making Language (TVML), a scripting language developed at NHK's Science and Technical Research Laboratory, to produce an entire TV program on a computer desktop as CG animation. We used the same scenario and direction for both Manzai-Robots and CG-animated characters. Figure 7 presents a display image of CG. We did not use human-like CG characters because our Manzai-Robots are not human-like robots. In our experiment, we present CG-animated characters on a 19-inch display. Participants were 18 people: the same as those for Experiment 1. After Experiment 1, participants watch CG-animated characters and also answer questionnaires. The questionnaires have a perspective that is the same as that of Experiment 1 (just change Manzai-Robots/News-robots into Manzai-Robots/CG in each questionnaire.

6.2.2 Results and discussion. The results presented in Figure 6(b) show only small differences in interest and understanding between Manzai-Robots and CG animations. We conducted a paired sample *t*-test to compare the Manzai-Robots and CG animations. Table VIII shows the results of the tests. For comprehensibility, the scores were $t(33) = 0.796$, $p = 0.43$; for interest, they were $t(32) = 0.978$, $p = 0.33$. The results indicate that participants did not judge a difference in interest and understandability between the Manzai-Robots and the CG animations. However, participants did experience a difference in familiarity with the Manzai-Robots being more familiar: $t(26) = 2.931$, $p = 0.0$.

The statistical results suggest that our techniques work well with animations and with robots in terms of creating interesting and understandable experiences of new stories. This outcome is not the one that we expected. We believe that these results indicate that humor, especially Manzai-style humor, is a compelling means of conveying information, even when robots and simple animated characters are presenting the humor. However, we draw no conclusions about familiarity because the animations use different characters than the Manzai-Robots. It is possible that our participants simply did not like the animations as well. In fact, 14 participants liked Manzai-Robots, and 4 participants liked CG animation.

6.3 Effectiveness of familiarity types for Manzai-Robots

The purpose of this research is to assess human feelings of familiarity and smoothness in human–robot communication using dialogue between robots based on a Manzai scenario. Two experiments reveal that the Manzai-Robots are familiar to humans. Familiarity of two types might be immediately felt familiarity and gradually felt familiarity. The former type includes familiarity of people who watch it for the first time.

	Mean	SD	*t*	DF	*p*
Understandable (M)	4.27	0.57	2.20	25	0.03
Understandable (N)	3.61	1.14			
Interesting (M)	3.88	0.58	5.82	27	0.00
Interesting (N)	2.27	1.01			
Familiar (M)	3.88	0.58	5.39	29	0.00
Familiar (N)	2.5	0.92			

Notes: M = Manzai-Robots; N = News-robots

Table VII. Result of *t*-test 1

Figure 7.
Display image of CG0animated characters

Table VIII.
Result of *t*-test 2

	Mean	SD	*t*	DF	*p*
Understandable (M)	4.27	0.57	0.79	33	0.43
Understandable (C)	4.11	0.67			
Interesting (M)	3.88	0.58	0.97	32	0.33
Interesting (C)	3.66	0.76			
Familiar (M)	3.88	0.58	2.93	26	0.00
Familiar (C)	3.05	1.05			

Notes: M = Manzai-Robots; C = CG animation

The latter are people who watch for some time and gradually feel increased familiarity. We consider that for familiarity in human–robot communication, the former type is more important than the latter type. We conducted an experiment to assess the effectiveness of types of familiarity of Manzai-Robots based on comparison of participants of two types who had never seen Manzai-Robots and those who had seen Manzai-Robots. In this section, we designate participants who have never watched Manzai-Robots as first-time participants and participants who had watched Manzai-Robots many times as frequent participants.

6.3.1 Conditions. Nine participants are first-time participants: another nine participants are frequent participants who watched Manzai-Robots three times and more during the prior half year. They watched the Manzai-Robots and answered the questionnaires. The questionnaires are the same as those used for Experiment 1. They responded to the questionnaire inquiries:

Q1. Do you feel familiarity with the Manzai-Robot behavior?

Q2. Do you feel familiarity with the Manzai-Robot emotions?

Q3. Do you feel familiarity with the Manzai-Robot voices?

Responses to the questionnaires were given on a five-point scale (5, highest; 3, middle; 1, lowest).

6.3.2 Results and discussion. The experiment results are presented in Figure 6(c) and Figure 6(d). Only a slight difference was found between first-time participants and frequent participants, but no statistically significant difference was found. This lack of difference suggests that our methods of creating news experiences work as well with new viewers as they do with frequent viewers. Regarding comparison of different comments, the first-time participants reported that it is difficult to understand what the robot is saying. That point persists as a problem of speech synthesis that we must consider the synthesis of speech. Results from Experiment 3 show that the Manzai-Robots lose familiarity only gradually. They are also attractive to people who have never watched Manzai-Robots.

7. Conclusion

We proposed a Manzai-Robot for which the interaction style is human-robot implicit communication based on an automatically generated scenario from Web news. We used the Manzai metaphor in our system: a traditional Japanese humorous comedy. Our generated Manzai scenario consists of snappy patter and a misunderstanding of dialogue based on the four kinds of gap of structure of funny points. Furthermore, we specifically examined the voice direction to assess the impression of robots. Via experimentation, we were able to propose a method by which the robots read the news by match voice for sentiment. We conducted experiments of three kinds to measure the following:

(1) the effectiveness of automatic generation of Manzai scenario for the robots;

(2) the effectiveness of Manzai-Robots as media; and

(3) the effectiveness of types of familiarity for Manzai-Robots.

Results show that the Manzai-Robots are effective for familiar and smooth human-robot communication.

As subjects of future studies, we expect to generate components of many kinds and increase the accuracy of those components. We will conduct experiments with people of various ages to measure the Manzai-Robot effects. Furthermore, we expect to create new features of Manzai-Robots such as changes of scenarios based on audience behavior and sentiment.

Notes

1. SONY AIBO, available at: www.sony.jp/products/Consumer/aibo/
2. HONDA ASIMO, available at: http://world.honda.com/ASIMO/
3. Softbank Pepper, available at: www.softbank.jp/robot/special/pepper/

References

Abe, T. (2006), "Dynamism in manzai", Bulletin of the Graduate Division of Letters, Arts and Sciences of Waseda University. III, Japanese Literature, Theatre and Film Arts, History of Fine Arts, Japanese Language and Culture, pp. 69-79.

Bouchet, F. and Sansonnet, J.P. (2009), "Subjectivity and cognitive biases modeling for a realistic and efficient assisting conversational agent", *Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology*, WI-IAT, IET, *Milan*, pp. 209-216.

GansoBakushoOu. (2008), *Introduction of Manzai*, Rittor Music, (ISBN: 4845615967).

Hayashi, K., Kanda, T., Miyashita, T., Ishiguro, H. and Hagita, N. (2005), "Robot manzai - robots conversation as a passive social medium", Humanoids, IEEE, Tsukuba, pp. 456-462.

Ishii, R., Nakano, Y.I. and Nishida, T. (2013), "Gaze awareness in conversational agents: estimating a user's conversational engagement from eye gaze", *Transactions on Interactive Intelligent Systems*, Vol. 3 No. 2, pp. 1-25.

Ishizaki, M. and Den, Y. (2001), "Computation and language: discourse and dialogue", University of Tokyo Press, Vol. 3, pp. 173–175.

Kanda, T., Ishiguro, H., Ono, T., Imai, M. and Nakatsu, R. (2002), "Development and evaluation of an interactive humanoid robot robovie", *Proceedings of the IEEE International Conference on Robotics and Automation*, *ICRA, IEEE*, pp. 1848-1855.

Khosla, R. and Chu, M.T. (2013), "Embodying care in matilda: an affective communication robot for emotional wellbeing of older people in Australian residential care facilities", *Transactions on Management Information Systems*, Vol. 4 No. 4, pp. 1-33.

Kitayama, D. and Sumiya, K. (2007), "Retrieving comparison articles using content elements order for news archives", IEEE Computer Society, Istambul, pp. 103-110.

Kumamoto, T. (2010), "Design of impression scales for assessing impressions of news articles", *Proceedings of the 15th International Conference on Database Systems for Advanced Applications*, *Heidelberg*, pp. 285-295.

Mashimo, R., Umetani, T., Kitamura, T. and Nadamoto, A. (2015), "Automatic generation of Japanese traditional funny scenario from web content based on web intelligence", *Proceedings of the 17th International Conference on Information Integration Web-based Applications & Services, iiWAS*, *ACM*, *New York, NY*, pp. 173-165.

Mashimo, R., Umetani, T., Kitamura, T. and Nadamoto, A. (2016), "Human-Robots implicit communication based on dialogue between robots using automatic generation of funny scenarios from web", *Proceedings of the 11th ACM/IEEE International Conference on Human Robot Interaction*, *IE, ACM/IEEE*, *Piscataway, NJ*, pp. 327-334.

Moriyama, T., Mori, S. and Ozawa, S. (2009), "A Synthesis method of emotional speech using subspace constraints in prosody", *Journal of Information Processing Society of Japan*, Vol. 50 No. 3, pp. 1181-1191.

Nakayama, K., Hara, T. and Nishio, S. (2007), "Wikipedia mining for an association web thesaurus construction", *Web Information Systems Engineering – WISE*, Springer, Heidelberg, Vol. 4831 pp. 322-334.

Park, S., Kang, S., Chung, S. and Song, J. (2009), "Newscube: delivering multiple aspects of news to mitigate media bias", *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, ACM, *New York, NY*, pp. 443-452.

Rae, I., Takayama, L. and Mutlu, B. (2013), "In-body experiences: embodiment, control, and trust in robot-mediated communication", *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI, ACM, *New York, NY*, pp. 1921-1930.

Shibata, T. (2010), "Integration of therapeutic robot, paro, into welfare systems", *Proceedings of the 28th Annual European Conference on Cognitive Ergonomics*, *ECCE, New York, NY*, pp. 3-3.

Shiomi, M., Sakamoto, D., Kanda, T., Ishi, C.T., Ishiguro, H. and Hagita, N. (2008), "A semi-autonomous communication robot: a field trial at a train station", *Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction*, *HRI, ACM*, *Amsterdam*, pp. 303-310.

Takizawa, O., Yanagita, M., Ito, A. and Isahara, H. (1996), "On computational processing of rhetorical expressions – puns, ironies and tautologies", *International Workshop on Computational Humor*, Enschede, pp. 39-52.

Tanaka, F., Movellan, J.R., Fortenberry, B. and Aisaka, K. (2006), "Daily hri evaluation at a classroom environment: reports from dance interaction experiments", *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-robot Interaction*, *HRI '06, ACM*, *New York, NY*, pp. 3-9.

Yamane, H. and Hagiwara, M. (2010), "A funny proverb generation system based on sukashi", *Proceedings of the 20th International Conference on Artificial Neural Networks: Part III, Lecture Notes in Computer Science*, ICANN '10, Springer, *Heidelberg*, Vol. 6354, pp. 368-377.

Yamaoka, F., Kanda, T., Ishiguro, H. and Hagita, N. (2006), "How contingent should a communication robot be?", *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction*, *HRI, ACM*, pp. 313-320.

Further reading

Hayashi, K., Sakamoto, D., Kanda, T., Shiomi, M., Koizumi, S., Ishiguro, H., Ogasawara, T. and Hagita, N. (2007), "Humanoid robots as a passive-social medium - a field experiment at a train station", *Proceedings of the 2rd ACM/IEEE International Conference on Human robot Interaction*, *HRI, ACM*, *Arlington, VA*, pp. 137-144.

Corresponding author

Akiyo Nadamoto can be contacted at: nadamoto@konan-u.ac.jp