The Robot Facilitating Conversation by Revoicing Keywords Learning from Active Conversations among Healthy Old Sisters

Tabito Kurosaka, Mihoko Otake

Graduate School of Engineering, Chiba University, Chiba City, Japan

{aaka2366, otake}@chiba-u.jp

Abstract

The purpose of this study is to develop a conversation assistive technology as a foundation of social networks in order to prevent cognitive decline and dementia of older adults.

We developed the system that supports users for switching attention by using a robot. In the group conversation, the robot takes 4 kinds of actions aiming at supporting switching attention by revoicing. In order that older adults can achieve the task, we counted number of switching attention used in natural conversation of real older adults, then set the number as a target number of switching attention system induce. By using the system we conducted experiments for younger adults and older adults.

As a result, the switching attention that the system induced reached the target number of switching attention in both younger adults and older adults. The system induced switching attention by not only revoice, but also other kinds of actions.

Introduction

Modern society faces dementia problem caused by rapid aging worldwide. In 2016, there are 46.8 million people worldwide living with dementia and it is anticipated to reach 131.5 million by year 2050.(World Alzheimer Report 2016) Although there is no certain way to prevent dementia yet, it is known that intensive social networks lower the risk for dementia. In this study, we aim to develop a conversation assistive technology as a foundation of social networks in order to prevent cognitive decline and dementia of older adults.

Japan is a super aging country and the number of elderly people increases every years. It is required to consider measure for the cognitive function problem. We employ healthy old sisters as role models who are known for their healthy longevities and active conversations. Analyzing their conversation it was confirmed that there were slight or no intervals when turn taking occurred (Otake,Yamaguchi 2013). Repeating keywords in the last sentences caused this turn taking. They were listening to the previous speeches carefully while they are preparing what to say next. In the conversation which has many turn takings, switching attention (Eric, Mei-Ching 2016) is used frequently. It is known that switching attention is one of the functions which decline with aging. If the system which supports to strengthen switching function is developed, it will be possible to prevent the function from getting lower.

It is estimated that healthy old sisters strengthen their cognitive function in everyday life. However it is difficult to take active conversations such as healthy old sisters. It will become easier to train cognitive function if we uses the technology that supports conversation.

Related Research

Robot using revoice

The robot that "revoice" human's utterances in a group conversation is studied in order to support conversation of human. In a collaborative learning, robot's revoice helped subjects solve the problem. They compared "Minimum revoicing condition and "Guiding- revoicing condition". When the situation was in "Minimum revoicing condition" , the quality of the answer was high. (Shirouzu, Miyake 2013)

The device facilitating conversation by interjection

There is a robot that induces active conversations by interjection. The robot was developed after analyzing the content and frequency of the interjection from real older adults' conversation. The robot interjection depends on the amount of speech. By interjecting to the speaker whose amount of the speech was lowest, it was estimated that amount of speech increased and the conversation got more active. (Shinchi 2016)

Attention Process Training

Sohlberg (Sohlberg 1986) categorized attentions into 4 groups as shown in Table 1. Table 2 shows the content of APT- II training. This is used to improve function of attention.

Table 1 Definition of Attention by Sohlberg

The name of Attention	The detail of attention
Sustained Attention	The function which maintain the reaction during a constant or repeated activity
Selective Attention	The function which maintain cognitive function requiring promotion and restraint of reaction towards several stimulation
Alternating Attention	The function which maintain flexibility which does different tasks alternatively
Divided Attention	The function which do different things at the same time.

Table 2 A part of APT-Ⅱ Training

The name of Attention	The content of training
Sustained attention	**Mental Math Activity** ・Participants hear the 4 numbers and answer twice number. (e.g.) Participants answer " 4 58 34 18" after they listened" 2 29 17 9" ・The length of time the participant spent and the ratio of correct answer in 10 pairs is recorded
Selective Attention	**Sustained Attention Activities with Distractor Noise** ・Participants do a "Sustained attention" task while distracter distracts participant. ・It is also good to do Puzzle. ・Distracter : Switching TV, radio on et al. ・The length of time the participant spent and the ratio of correct answer is recorded

Alternating attention	**Alternating Alphabet Exercise**
	・Participant fill the "before" or "after " block in alphabets line.
	・The place of "before" or "after" block changes in constant time
	  (e.g.) If the block is "before"
	    Answer "R" if it is "S","A" if it is "B"
	    If the block is "after"
	    Answer "T" if it is "S","C" if it is "B"
	・The length of time the participant spent and the ratio of correct answer is recorded

Divided Attention	**Read and Scan Task** ・Participants deletes target letter("-ing","snd") with reading articles. ・Participants have to understand the content of the article. ・After training, participants asked a question about article. ・The length of time the participant spent, the quality of deleting target, the level of understanding about article is recorded

Method

We made a robot to attract subjects' attention and conducted an experiment. In order to confirm that older adults can achieve the task, the system was designed based on the real older adults' conversation analysis. We counted number of switching attention used in natural conversations of real older adults.

Experiment

The subjects were asked to talk to each other for 3 minutes using the designated topic in Fig 1. The designated topic was shown to subjects in a topic sheet.

During the experiment, the subjects were given instruction to pay attention to the robot when the robot actions, although the actions of robot are unpredictable. Besides, the subjects were asked to continue talking while switching their attention.

The layout of experiment is shown in Fig 2. 2 cameras are used to record the subject's behavior. The microphone connected to a computer is to recognize subject's speech. After the recognition process, the processed signals are sent to robot via computer so that the robot can take actions. In order for subject to watch robot and their experimental partner equally, the robot is set in between two subjects, and the angle between the subject and the robot is 60 degrees. The experiment was conducted with pairs of 1 male and 1 female. The subjects were 6 younger adults (3 Males, 3Females, average age: 23) and 5 older adults. (2 Males, 3Females, average age: 78).

Please talk to each other for 3 minutes after the starting call of the robot.
    When the robot talks to you, please watch the robot or reply.

Kyuri (Cucumber)
Tamanegi (Onion)
Okura (Gumbo)
Kyabetsu (cabbage)
Tomato
Yamaimo (Japanese yam)
Horensou (Spinach)
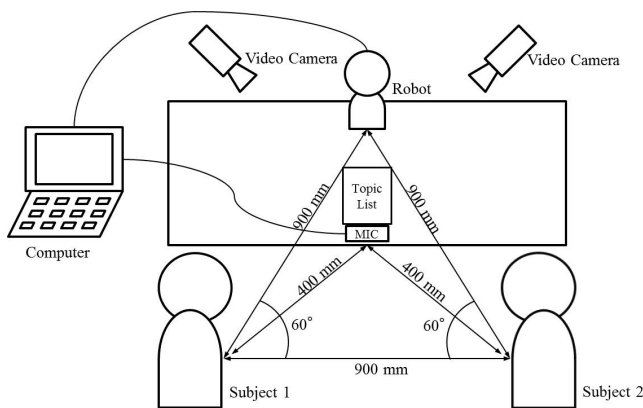Shimeji (Shimeji mashroom)

Fig.1 Topic List

Fig 2 Experiment Overview

Fig 4 Number of times healthy older adults' switching attention
in 223 seconds conversation

The analysis of the conversation of healthy old sisters

We analyzed conversation data of 3 Japanese healthy old adults in Fig 3. They are sisters and known for their healthy living by having active group conversations in everyday life. We counted how many times they used switching attention during 223(sec) times of free conversation. Fig 4 shows number of switching attention depends on each old adult. If A looks at B when B talks to A while A talks to C, action of A was counted as switching attention. In 223(sec) times of conversation, A switched attention 6 times, B switched attention 10 times, C switched attention 12 times. We calculated frequency of switching attention for each older adult, then calculated the average. We foundaverage number of switching attention in 3 minutes was about 7 times.

Fig 3 Japanese healthy old sisters who enjoy active
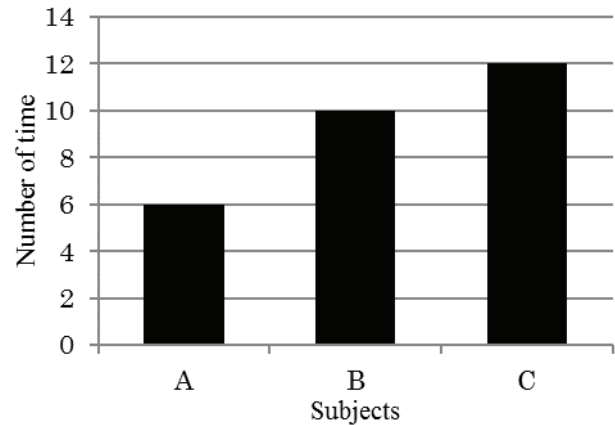conversations in everyday life

System

Fig 5 is the flow chart of the system used in experiment. At the blocks of "Ask a question", "Interjection", " Revoice", the system sends a signal to the robot and the robot takes action. The servo motor in robot's neck rotates and the speaker in the robot reproduces registered scripts. Table 3 is the scripts and the condition which cause actions. We aimed to cause subject's switching attention with using "Revoice" and "Ask a question". The system ends 3 minutes after "Start" block.
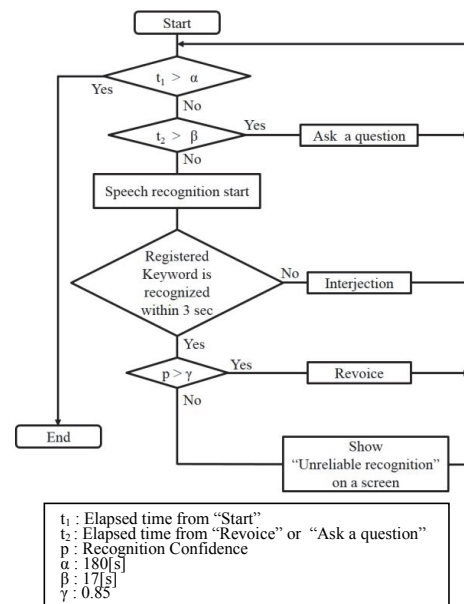

Fig 5 Flowchart of the system

Here is the detail of 4 block, "Revoice", "Ask a Question", "Interjection", "Show 'Unreliable recognition' on a screen".

"Revoice" block

If the elapsed time from the last time "Revoice" or "Ask a Question" within 17 sec, speech recognition starts. And the system reproduces registered scripts in Table 3 when speech recognition engine recognized registered keywords within 3 seconds. The scripts are real scripts in a free conversation data about vegetables by healthy old sisters except for registered keywords "Oishii(Delicious)", "Umai(Delicious)", "Kirai(Hate)", "Nigate(Bad)", and reproduced scripts "Oishiidesune(It's delicious.)", "Kirainandesune(You hate it.)". We added those additional keywords because those appear frequently in conversation about vegetables. The total time of conversations was 20 minutes to pick up keywords to revoice. There are 2 kind of reproduced scripts, one is just a vegetable name and the other is vegetable name and other information. This is because the keyword consisting vegetable name and other information induce subject's attention better.

"Ask a Question" block

Considering the situation when there are no registered key words in conversation, we added "Ask a question" block to the system. This block is executed after 17 seconds from the last time "Revoice" or "Ask a Question" block is executed.

In "Ask a question" block, the system reproduces questionnaire in table 3. The expected situation is as follows. Robot asks a question "How about an onion?". Subject says "I like onion". Then the system recognizes a registered keyword "onion" in the subject's speech, and reproduces "It's also good to fry onion". By "Ask a question" block, the possibility of registered keywords being used increase. If the registered keywords are used, the system steps on "Revoice" block.

"Interjection" block

When both subjects are silent for 3 seconds or there are no recognition words within 3 seconds, the system steps on "Interjection" block. Because interjection represents a signal "Keep the conversation" (Schegloff 1982). The system prompts subject's utterance by reproducing "Interjection". Real scripts and ratio of appearance of healthy old sister's interjection (Shinchi 2016) is reflected in "Interjection" block.

"Show 'Unreliable recognition' on a screen" block

If the system recognized a registered keyword but the confidence of recognition is less than β, the system show "Unreliable recognition" on the console display, then the

process returns to the head of loop. Robot doesn't take action in this block.

Table 3 Reproduced scripts and conditions

	The condition	Reproduced Scripts
Revoice	**Registered Keywords** below is pronounced	
	Kyuuri (Cucumber)	Kyuuri (Cucumber)
	Kyabetsu (Cabbage)	Kyabetsu (Cabbage)
	Shimeji (Mushoom)	Shimeji (Mushroom)
	Tomato	Tomato
	Yamaimo (Japanese yam)	Yamaimo wa tororo, oishiine. (Yamaimo is good to grate.)
	Tamanegi (Onion)	Tamanegi wa itametemo oishiine. (Tamangi is also good to fry.)
	Okura (Gumbo)	Okura no hana wa monosugoku kireinano. (Okura's flower is so beautful.)
	Horensou (Spinach)	Horensou wa ohitashi ne. (Horensou is good to boil.)
	Oishii (Delicious)	Oishiidesune. (It's delicious.)
	Umai (Delicious)	
	Kirai (Hate)	Kirai nanndesu ne. (You hate it.)
	Nigate (Bad)	
Question	β [sec] elapsed after "Revoice" or "Question"	"Vegetable Name" wa doudesuka ? (How about "Vegetable Name"?) (Vegetable Name: Kyuri/Kyabetsu/Shimeji/ Tomto/Yamaimo/Tamanegi / Okura/ Horensou)
Interjection	When the system doesn't recognize **Registered Keywords** within 3 seconds	Un(Yes)
		Sou Sou(Yes Yes)
		Ah(Ah)
		Huun(Ah)
		Honntou(Really?)
Starting utterance	At the first of the experiment	Soredewa ohanashi wo hajimete kudasai, douzo. (Please start to talk.)

Robot

Table 4 shows the Specification of the robot used in this research. The neck part of the robot moves by servo motor. Fig 6 shows the appearance of the robot. We scanned the

shape of the body of one of the healthy old sisters, then printed with 3D printer by permission of her.

Table 4 Specification of the robot

Size	x130 y120 z240 (mm)
Weight	400g
Body Material	ABS
Degrees of Freedom	1DOF(Neck)
Controller	Personal Computer
Power	DC 9V 2.5A

Front view   *Side View*
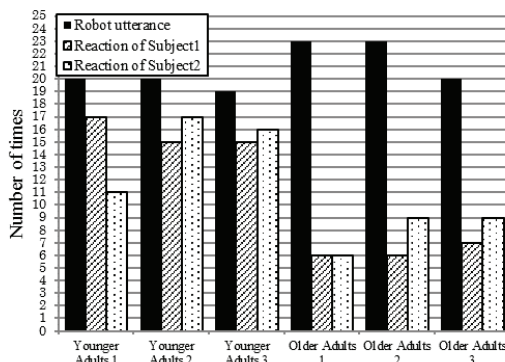*Fig 6 Overview of the robot*

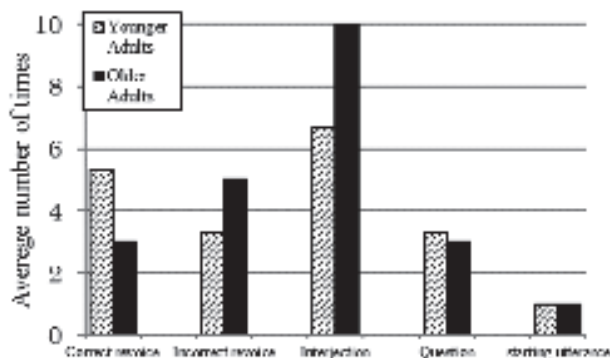*Fig 7 Number of times robot actioned and subjects reacted*
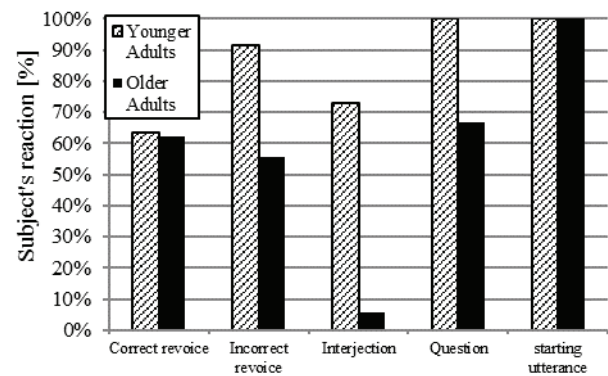

Fig 8 Average number of robot action

Fig 9 Average of subject reaction rate to robot utterance

Results

Fig 7 shows number of times robot actioned and subjects reacted. The robot's actions and the subject's reactions watching the robot or replying to the robot when the subject was looking at other object counted in the movie recorded in an experiment. When the subject reacted to the action of the robot while the subject watched a robot, the reaction of the subject was not counted because the attention didn't switch from the other subject to the robot. Younger Adults 1, Younger Adults 2, Younger Adults 3 are the pairs of younger adults, and Older Adults 1, Older Adults 2, Older Adults 3 are the pairs of older adults. The average number of the reaction of Younger Adults was 15.2 times, and the average number of the reaction of Older Adults was 7.2 times. Even though the numbers of robot's action in Older Adults experiment were more than that of Younger Adults experiment, the numbers of reaction of Older Adults were less than that of Younger Adults.

Fig 8 shows the average number of robot action caused by subjects' utterances. "Correct revoice" was the reproduced scripts when the system recognized registered keywords correctly, and "Incorrect revoice" was the reproduced scripts when the system recognized the registered keywords incorrectly.

The sum of average numbers of "Correct revoice" and "Question" was about 9 times in Younger Adults experiment and was about 6 times in Older Adults experiment. The average numbers of "Interjection" action were largest in both Younger Adults and Older Adults.

Fig 9 indicates average of subject reaction rate to each action of the robot. In Younger Adults, the ratio of reaction to "starting utterance" and "Question" was 100(%). The reaction rate to "Incorrect revoice" was the second highest of subject reaction, and the reaction rate to "Interjection" was the third highest of subjects reaction. In Older Adults, reaction rate to "starting utterance" was also 100(%). The

reaction rates to "Correct Revoice", "Incorrect voice", "Question" were about 60(%).

Discussion

In Fig 7, average number of times subjects reacted to a robot in 3 minutes was 15.2 times among Younger Adults and 7.2 times among Older Adults. The occurrence of the subjects' reaction to the robot while they are talking means that they switched attention to the robot. The target number of switching attention was 7 times in 3 minutes. Thus subjects' switching attentions induced by the system reached the target number of switching attention for both Younger Adults and Older Adults. Even though the number of robot's action to Older Adults was more than that of Younger Adults, the reaction of Older Adults was less than that of Younger Adults. This was because in Fig 8, even though average number of times robot injected to Older Adults was much more than that to Younger Adults, average reaction rate of Older Adult's to the robot's "Interjection" action was much less than that of Younger Adults shown in Fig 9. There was a reaction rate difference between Younger Adults and Older Adults.

As Fig 9 shows, the reaction rates of both Younger Adults and Older Adults were highest in "starting utterance". Because "starting utterance" was the first action in an experiment, this action was highly reacted.

In Younger Adults experiment, average of subjects reaction rate to "Incorrect Revoice", "Question" was higher than "Correct Revoice" and "Interjection". When the system steps on "Incorrect Revoice" and "Question", it was observed subjects stopped conversation, and reacted to the robot. This indicates that the subject tends to react to the robot when the robot reproduces scripts not matching the content of the conversation rather than scripts matching the content of the conversation.

In Older Adults experiment, average of subjects reaction rate to "Correct Revoice", "Question", "Incorrect Revoice" was higher than "Interjection". As for "Correct revoice" rate, it consists reaction rate of Older Adults 1, Older Adults 2 and Older Adult 3. However there is a possibility subject's reaction rate of Older Adults 3 was exceptional and the number should be eliminated. In an experiment in Older Adults 3, the reaction rate was 100 (%) because there was only 1 "Correct Revoice" to which subject reacted. It is not valid value because the action was only once. If the reaction rate of Older Adults 3 is eliminated, average of subjects' reaction rate of Older Adults was 43(%). The reaction rate of "Incorrect revoice" and "Question" is higher than that of "correct revoice" and "Interjection", which is as same as Younger Adults. The reaction rate of "Interjection" was much less than that of Younger Adults. Older

Adults continued to talk without reaction to the robot when robot interjected, while Younger Adults reacted.

In Fig 8, the system induced "Interjection" the most in both Younger Adults and Older Adults. As Fig 4 shows, the system recognizes sound for 3 seconds. However there were utterances over 3 seconds in an experiment. Thus there were many cases that the system induced "Interjection" while subjects speaking. Except for "Interjection", there were many "Correct revoice" in younger adult experiment, "Incorrect revoice" in older adult experiment. Younger Adults' speeches were clearer and faster than those of Older Adults. Therefore, the system could not recognize accurately for Older Adults' speech.

Conclusion

We developed a system that induces switching attention in a group conversation. The system was aimed to induce switching attention by using revoice and question that prompts revoice. The number of switching attention used in a conversation of real older adults was counted, then we set that number as a target number of switching attention the system induce.

The conclusion of this research is summarized as:

・The number of times system induced subject's attention reached target number.
・Switching attention was induced by not only "Revoice" and "Question", but also other actions.
・There is a possibility that utterance and question which is no relation to the conversation induce switching attention.

In this research, the system induce switching attention by not only "revoice" and "question", but also other actions. And the switching attention was counted by subjective evaluation. Future work of this research includes evaluating actions which is no relation to the conversation and recompose the design of the system. In order the system recognizes speech stably, we will adjust the length of recognition time. Then we conduct experiment with valid number of subjects.

References

Otake M., Yamaguchi K. 2013. Analysis of Overlap during Group Conversation of Active Older Adults, *Proceeding of the 27th Annual Conference of the Japanese Society for Artificial Intelligence*: 3C1-IOS-1b-5

Schegloff, Emanuel and Harvey Sacks.1982. Discourse as an interactional achievement:: Some use of "UH-HUH" and other things that come between sentences. In Deborah Tannen (ed.), *Georgetown University Round Table on Languages and Linguistics, Analyzing discourse*: *Text and talk*, 71-93.Washington DC: Georgetown University Press

Shinchi, A. 2016.Analysis of active conversation and its application to the development a support device for group conversation, p 30, Graduate school of Engineering, Chiba university, Chiba city, Chiba prefecture, Japan.

Shirouzu, H., and Miyake, N., 2013. Effects of Robots' Revoicing on Preparation for Future Learning. In Rummel, N.,Kapur,M.,Nathan,M., and Putambekar (Eds.) *In proceedings of CSCL 2013 Conference, Wisconsin, Vol.1 438-445*

Sohlberg MM, Mateer CA. 1986. Attention Process Training for neuropsychological Research and Development, Washington DC.

Sohlberg MM, McLaughlin KA et al.1993. The Manual for Attention Process Training-Ⅱ. A Program to Address Attentional Deficits for persons with Mild Cognitive Dysfunction. AFNRD, WA.

World Alzheimer Report 2016. Martin Prince, Ms Adelina Comas-Herrera, Martin Knapp, Maëlenn Guerchet, Ms Maria Karagiannidou.2016.World Alzheimer Report 2016, The Global Observatory for Ageing and Dementia Care, King's College London and the Personal Social Services Research Unit (PSSRU), London School of Economics and Political Science.