International Journal of Service and Knowledge Management
International Institute of Applied Informatics
2018, Vol.2, No.1, P.19 - 30

Topic Visualization System for Unfamiliar Couples in Face-to-Face Conversations

Yoko Nishihara *, Kodai Yoshimatsu *,
Ryosuke Yamanishi*, Satoshi Miyake*

Abstract

This paper proposes a topic visualization system for unfamiliar couples in face-to-face conversations. Some of people are weak at making face-to-face conversations with people for the first time. It is so hard to find good topics for conversations. If any topic can not be found, a non-speech section will come to a conversation and will make them be more awkward. The proposed system visualizes topics to users. As inputting the user's profile data into the system, conversation topics are visualized by referring the input. The proposed system recognizes their speeches in conversations, and detects non-speech sections in their conversation. If a non-speech section is detected, the system switches topics to new ones and presents the new topics to the users. They can continue their conversations if non-speech sections come more than one time. We experimented the efficiency of the proposed system in decreasing non-speech time in conversations.

Keywords: Face-to-face conversation, Unfamiliar couples, Topic visualization, Silent time

1 Introduction

Conversation support methods have been widely researched in many academic and business areas not only for human-to-robot conversations [3, 14, 16, 18] but also for human-to-human conversations [13]. We also have researched human-to-human conversation methods [15]. In human-to-human conversations, they often use Web services such as short message services and chatting services. The services enable users to send messages to their friends as soon as they want. They never wait for response messages with holding their mobile phones until the response messages come. When the responses come to their mobile phones with notifications, they would read those. They can write and read their messages whenever they would like to use the services. Such services have set people free from the pressure to communicate with others.

With the spreading of the Web services, people have less time of making face-to-face conversations with others [10]. People used to enjoy conversations while having lunches and dinners before inventing of the Web services. Nowadays, in their lunches and dinners, they use their own mobile phones to make conversations with others who are far from them. They do not make conversations with their friends at the same table. People, especially young people, often are afraid that they are weak in making face-to-face conversations. Though they can talk with their friends and members of their own family, they

* Ritsumeikan University

can not start to talk with others for the first time. Even if they start to talk, they can not continue their talks. However, they can talk with anyone on the Web services. Why is it hard to start and continue their conversations in the face-to-face situation? The answer for the question should be that they do not find good topics for their conversations. Without finding any topics to talk with, they can not speak anything and then a non-speech section comes. A non-speech section is uncomfortable for people who feel their weakness in their ability to make conversations. On the other hand, they have enough time to find good topics in conversations on the Web services. Therefore, they may not feel uncomfortable in conversations on the Web services. People do not like non-speech sections. To decrease the non-speech time in the face-to-face conversations, some methods to find good topics are required.

In our previous paper, we proposed a conversation support system [17]. The previous system supports face-to-face conversations of unfamiliar couples. The users input their profile data into the system. Then, the system selects and presents topics that are common to talk for both of them. The system uses words of hobbies in their profile data to selects topics. The system selects words related to the words of hobbies based on a thesaurus in the topic selection and presents the words as the recommended topics for the conversation. We confirmed that the system could decrease the non-speech sections in conversations. The previous system presents topics only one time. The users can avoid the first non-speech section by using the presented topics, however, they may not avoid the second and later non-speech sections. When a non-speech section is detected, the topics should be switched to new ones, and the new topics should be presented to the users.

We do not have any findings of the timing and what the non-speech section should be for switching topics. In this paper, we conduct preliminary experiments to find the best non-speech section for switching topics. Also, we propose a new system that supports conversations by switching topics when a non-speech section comes.

People often wonder the timing to speak or not in their conversations. The behavior is called as the conversational inhibition, which has been studied by many researchers[2, 4, 6, 7, 8]. One of these findings showed that talking with an unfamiliar person had bigger number than talking with a friend for the conditions getting conversational inhibitions[9]. In this paper, we try to support unfamiliar couples who tend to get conversational inhibitions by visualizing common topics. The expected outcome of our study is that the visualization of topics will support the users to continue their conversations.

2 Preliminary experiments

We conducted preliminary experiments to determine the time for switching topics for users in the proposed system. Our previous system [17] does not switch topics in a conversation. The previous system visualizes the topics on the interface at the beginning of a conversation. System users in a conversation can avoid the first non-speech section by using one of the presented topics by selecting as a next topic. However, the users may not avoid the second or later non-speech section in their conversations. The visualized topics should be switched every timing that a non-speech section comes, some of the non-speech section works for their consideration, though. Then how long time should be for switching the topics? How many seconds should the system keep to visualize the same topic in a non-speech section?

2.1 Experimental procedures

The procedures of the preliminary experiments will be shown in the following paragraph.

1. The experimenter (who was the second author) made pairs of participants. Each pair had two participants. The number of pairs was 10. The experimenter divided the 10 pairs into two groups. The two groups were called as an unfamiliar group, and as a familiar group.

2. The participants of the preliminary experiments were asked to fill their profiles: account name, gender, living place, name, age, blood type, birthday, part-time job, favorite actor/actress, and hobby.

3. Each pair in the two groups was asked to make a conversation. All pairs in the two groups used the previous system in their conversations. All the conversations were recorded with a voice recorder.

4. After five minutes from the beginning of the conversation, the experimenter asked each pair to stop their conversation.

Each pair in the two groups consisted of a male and a female participant. The participants were undergraduate students majoring information science and engineering. The range of their age was between 21 to 23 years old.

The previous system visualizes topics on the interface as shown in Fig. 1. The topics are selected referring the relations between their hobbies in their profiles.

When dividing the pairs into the groups, the experimenter asked the pairs "Did you two make a conversation for more than five minutes before?" If a pair answered YES, the pair was classified into the familiar group, and if no, the pair was classified into the unfamiliar group.

We asked the participants to ask some questions and give opinions relating to their profiles to the partner in the pair. The participants could select the next topic by deriving a topic relating to their profiles. The participants were asked to speak to microphones for their voice data to be recorded. The conversation time for each couple was set to five minutes because unfamiliar couples often feel hard to continue their conversations for the first five minutes. If they get over the first five minutes, they may be able to continue their conversations more. Therefore, we select the time for each conversation.

2.2 How to measure non-speech section in a conversation

We measured non-speech sections in a conversation. The non-speech section denotes a section in which both of the participants in a pair do not speak anything. We used Audacity, which is free audio software [1], to obtain volume data of recorded voice data every second. We defined a section where the volume level is higher than or equal to 0.01 as a speech section. The non-speech section is defined as a section where the volume level is less than 0.01.

We removed waiting sections for answers to questions from the set of the non-speech sections. The rest of the set was assumed as the non-speech sections.

Each of the non-speech sections had the time information of the beginning and the end of the section. We calculated the difference between the times for the beginning and the end. The difference was used as the time for a non-speech section.

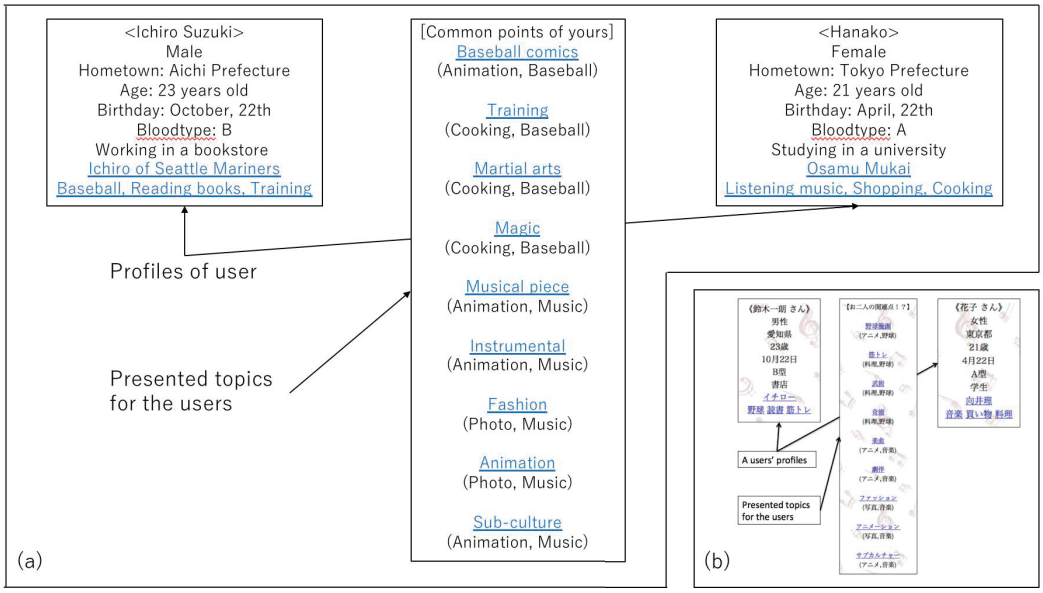

Figure 1: The interface of the previous our system supporting conversations of couples for the first time. (a) is translated information of (b) in English . We used the interface (b) in experiments. Their profiles are shown on the left and the right side. The topics are shown in the center. The topics are selected referring the relations between hobbies of the two profiles. In this example, the hobbies of a user on the left side are "baseball," "reading books," and "training." The hobbies on the right side are "listening to music," "shopping," and "cooking." The topics in the center are "baseball comics," "training," "martial arts," "magic," "musical piece," "instrumental," "fashion," "animation," and "sub-culture."

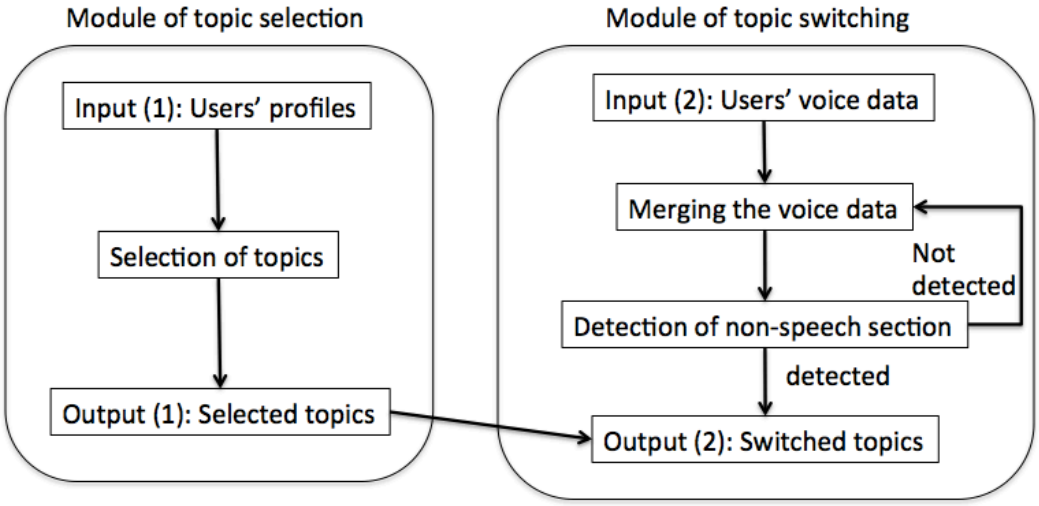

Figure 2: Outline of the proposed system. The proposed system consists of two modules: a module for topic selection and a module for topic switching.

2.3 Experimental Results

The all pairs in the groups firstly asked questions for their profiles each other. Then they started to talk other topics. The average time of non-speech sections for the familiar group was 1.5 seconds though the one for the unfamiliar group was 2.5 seconds. The average time for the familiar group was significantly higher than the one for the unfamiliar group (Student's t-test, t=2.7, p=0.02 $<$ 0.05). It seemed that the pairs in the unfamiliar group had trouble with finding the topics for their conversations. If the topics were presented many times to continue their conversations, they may decrease the non-speech sections in their conversations.

According to the results, we determined to develop a new system that equips a function to switch topics if a non-speech section continues for 2.5 seconds.

3 Proposed visualization system switching topics every non-speech section

Fig. 2 shows the general idea of the proposed system. The proposed system consists of two modules: a module for topic selection and topic switching. Each module has each input and output.

The proposed system is developed to be used by two users. The positions of the users, their microphones, the interface, and the proposed system are shown in Fig. 3. The users have their seat next to each other. They speak to their microphones. The voice data is sent to the proposed system. They see topics visualized on the interface. Fig. 4 shows an example of topics presented on the interface.

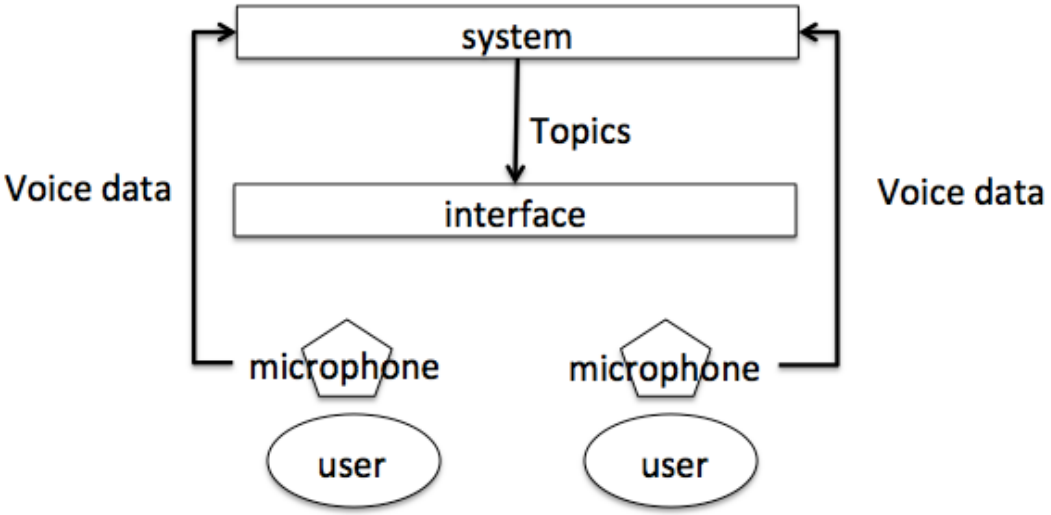

Figure 3: Positions of the users, the microphones, the interface, and the proposed system.

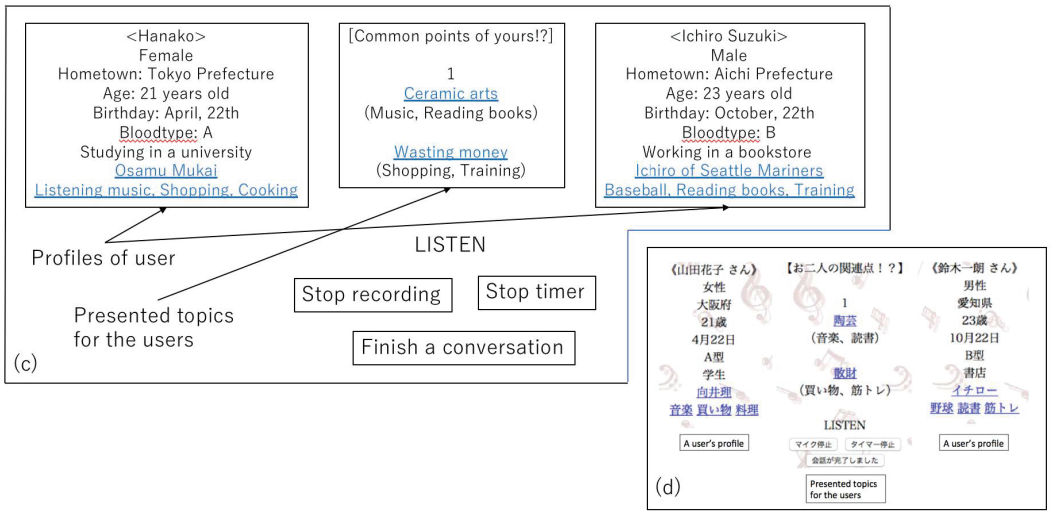

Figure 4: Example of topics visualized on the interface. (c) is translated information of (d) in English . We used the interface (d) in experiments. The users' profiles are shown in the left and right side. The topics are shown in the center. The topics are switched every timing non-speech section occurs. In this example, the hobbies of the user on the left side are "listening music," "shopping," and "cooking." The hobbies on the right side are "baseball," "reading books," and "training." The presented topics in the center are "ceramic arts" and "wasting money."

3.1 Module for topic selection

The input data of the module for topic selection is two users' profiles. The proposed system is used by two users simultaneously. The module selects topics referring their hobbies in the profile data and put the topics as the output.

3.1.1 Input: Users' profiles

The system users input their profile data. Each user inputs account name, gender, living place name, age, blood type, birthday, part-time job, favorite actor/actress, and hobby. The profile data is the same as those which were used in our previous system [17].

3.1.2 Topic selection

In the system, Word2Vec is used to select topics [5, 12]. We use Japanese Wikipedia's data in the learning on Word2Vec. After the learning, the module inputs each word of hobby on the profile to Word2Vec. Word2Vec outputs each vector of words for the input word, which is hobby.

The module selects words that are common to two vectors: a vector from a user and a vector from the other user. The selected words are ranked with the index number of cosine similarity in a vector. Then, the selected and ranked words are used as the topics in the module for topic switching. The module does not consider a topic in the current conversation. The authors will try to find the limitation of the module that consider only the users' profiles.

3.2 Module for topic switching

The input data of the module for topic switching is the users' voice data. The voice data is obtained through two microphones for the two users. The module merges the voice data of the two users. If a non-speech section is detected, the module decides that the users should switch the topic of the conversation, and outputs the next topics based on the output of topic selection module.

3.2.1 Merging of the users' voice data

The voice data of each user is sent to the module separately. The module recognizes each voice data by using a speech recognition machine, which is Julius [11] in this paper. Recognized voice data is output as text data. The module merges the text data in time order. The merged data is saved as a log file with time information.

In speech recognition, Julius also records three states: the beginning of speaking, the end of speaking, and waiting period for the speaking. The waiting period for the speaking is marked just after the end of speaking. The three states are also written in the log file with time information.

3.2.2 Detection of non-speech section

The module detects non-speech sections by using the three states recorded by Julius. Fig. 5 shows the procedure to detect the non-speech section. The module starts to detect a non-speech section when the states of both of the two users are waiting for speaking. The module stops detecting the non-speech section when the state of either user changes to the beginning

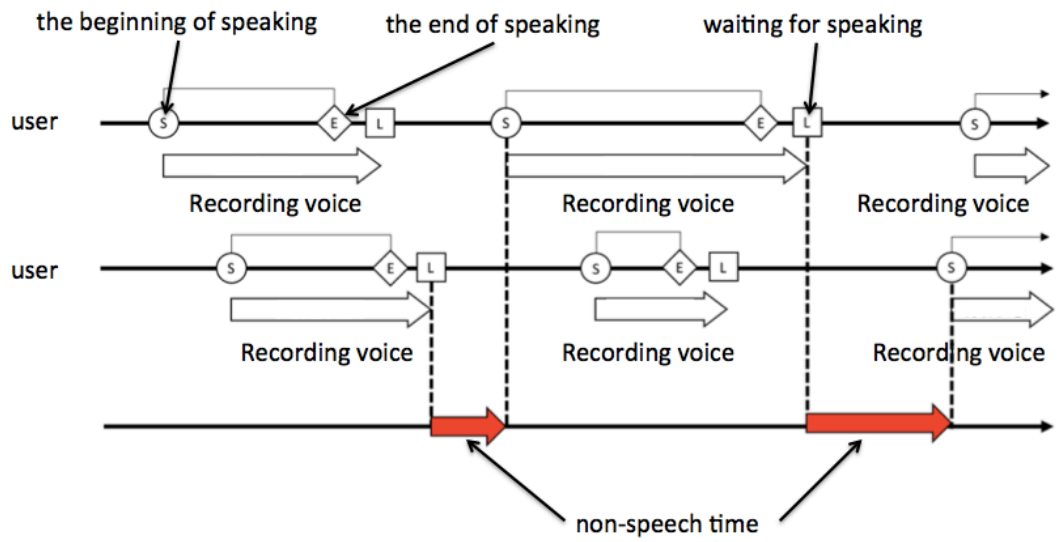

Figure 5: The procedure of detecting non-speech sections.

of speaking. If either user does not start to speak in 2.5 seconds from the beginning of the detection, the module switches topics and presents them to the users on the interface.

4 Evaluation experiments

We conducted the evaluation experiments with the proposed system. The purpose of the experiments was to verify the efficiency of the proposed system in decreasing non-speech sections in the conversations.

4.1 Experimental procedures

The procedures of the evaluation experiments are shown in the following paragraphs.

1. The experimenter (who was the second author) made pairs of participants. Each pair had two participants. The number of pairs was 18. The experimenter divided the 18 pairs into three groups. The groups were a proposed group, a comparative group, and a baseline group.

2. The participants of the evaluation experiments were asked to fill their profiles: account name, gender, living place name, age, blood type, birthday, part-time job, favorite actor/actress, and hobby.

3. Each pair in the three groups was asked to make a conversation. The pairs of the proposed group used the proposed system in their conversations. The pairs of the comparative group used a comparative system in their conversations. The pairs of the baseline group used our previous system in their conversations. All the conversations were recorded with a voice recorder.

4. After five minutes from the beginning of the conversation, the experimenter asked each pair to stop their conversation.

Each pair in the three groups consisted of a male and a female participant who met for the first time. The participants were undergraduate students majoring information science and engineering. The range of their age was between 21 to 23 years old.

As mentioned above, we prepared the three groups for the experiments. The proposed group used the proposed system. The baseline group used our previous system. The comparative group used a comparative system that switched topics if a non-speech section continues for 1.5 seconds. We prepared the comparative system to verify the appropriate interval for switching topics. The interval of the comparative was determined based on the results of the preliminary experiments.

We asked the participants not to give questions or opinions that are not related to the information on their profiles. The participants could select the next topics by deriving a topic relating to their profiles. The participants were asked to speak to microphones so that their voice data could be recorded.

We measured non-speech sections in their conversations. We supposed that the non-speech sections were decreased in the proposed group. The procedure to measure the non-speech sections was the same as those in the preliminary experiments.

4.2 Experimental results

Table 1 shows the obtained non-speech sections: the number of non-speech sections, the total time of the non-speech sections, and the averaged time of the non-speech sections for each group, respectively.

The average time of the non-speech sections was 7.9 seconds in the conversations with the proposed system, though 11.2 seconds with the comparative and 9.4 seconds with the baseline system. The shortest time of the non-speech sections was confirmed by using the proposed system.

5 Discussions

The experimental results showed that the shortest time of the non-speech sections was confirmed with the proposed system. In comparing the proposed system with the baseline system, the function of switching the topics every non-speech section worked efficiently to find the next topics in the conversations.

The averaged numbers of non-speech sections were the same between the proposed system and the baseline system. In contrast, the averaged time of non-speech sections were different between them. It indicates that the participants of the proposed system could select topics instantly and continue the conversations. The proposed system shows a few topics on the interface while the baseline system shows all topics on the interface. Therefore, the participants could select topics instantly.

In comparing the proposed system with the comparative system, the time of the non-speech sections with the comparative system was higher than the one with the proposed system. The comparative system switched the topic to the next topics every timing when 1.5 seconds of non-speech section occurred. The switching interval of the comparative system was shorter than that of the proposed system. In the comparative system, the interval might be too short for the users to check up the next topics shown on the interface. The topic might be switched interrupting the checking of the next topics. The participants could not find the topics to continue their conversations because of the fast switching of the topics.

Table 1: Number and time of the non-speech sections for each system.

Proposed system			
Pair	# of the non-speech sections	total time	averaged time
1	4	7.36	1.84
2	2	4.57	2.28
3	5	13.72	2.74
4	3	4.27	1.42
5	5	13.99	2.80
6	2	3.81	1.91
Avg.	3.5	7.9	2.2
Comparative system			
Pair	# of the non-speech sections	total time	averaged time
1	4	8.86	2.22
2	8	19.52	3.23
3	4	7.42	3.54
4	5	9.83	3.59
5	4	9.08	4.05
6	6	12.34	3.44
Avg.	5.2	11.2	3.3
Baseline system			
Pair	# of the non-speech sections	total time	averaged time
1	8	24.67	3.08
2	3	6.87	2.29
3	2	7.76	3.88
4	2	5.32	2.66
5	1	0.89	0.89
6	5	10.99	2.20
Avg.	3.5	9.4	2.5

The experimental results proved that the proposed system can reduce non-speech time more than the baseline system. We need to improve the proposed system as the future work that supports users for enjoyable conversations. The proposed system shows topics that are selected by evaluation of relations between users' profiles. We will improve the selection method of topics in order to realize the enjoyable conversations. It may be good if a keyword in a profile is more favorite one for a user, the system evaluates the keyword as an important one.

6 Conclusions

This paper proposed a system that supports conversations of unfamiliar couples. The system visualizes topics for conversations to users. Their voice data in conversations was recognized, and non-speech sections were detected. If a non-speech section was detected, the system switched topics to new ones and presented the new topics to the users. The users could continue their conversations if non-speech sections come more than one time.

We experimented with three systems: the proposed system, a comparative system and a baseline system. The baseline system visualized topics for only one time. The comparative system visualized topics every timing when a non-speech section for 1.5 seconds occurred. The interval of switching topics in the comparative was shorter than those in the proposed system. Experimental results showed that the time of non-speech sections with the proposed system was shortest among three different systems. We confirmed that the proposed system contributed to decrease the time of the non-speech sections by switching topics every non-speech section of 2.5 seconds.

As the future work, we will survey the effectiveness of the proposed system as a practicing tool for making enjoyable conversations.

Acknowledgment

This work was partially supported by JSPS KAKENHI Grant Number 16K00307.

References

[1] http://www.audacityteam.org/ (accessed on 16th March 2017)

[2] Penelope Brown and Stephen C. Levinson, Politeness: Some universals in language usage, New York: Cambridge University Press, 1987.

[3] Dagoberto Cruz-Sandoval, Friederike Eyssel, Jesus Favela, and Eduardo B. Sandoval, Towards a Conversational Corpus for Human-Robot Conversations, Towards a Conversational Corpus for Human-Robot Conversations, pp.99–107, 2017.

[4] John Augustine Daly and James C. McCroskey, Avoiding communication: shyness, reticence, and communication apprehension, Sage Publications, 1984.

[5] https://deeplearning4j.org/ja/ja-word2vec (accessed on 16th March 2017)

[6] William T. O'Donohue, Handbook of Psychological Skills Training: Clinical Techniques and Applications, Allyn and Bacon, 1995.

[7] Jackie Gnepp and Debra L. Hess, Children's understanding of verbal and facial display rules, Developmental Psychology, Vol. 22, pp.103–108, 1986.

[8] Daniel Goleman, Emotional Intelligence, New York: Bantam Books, 1995.

[9] Miho Hatanaka, Decision-making processes underlying conversational inhibition, Japanese Journal of Social Psychology, Vol.21, No.3, pp.187–200, 2006. (in Japanese)

[10] Emiko Isogai and Takamasa Suzuki, DENSO Technical Communication Education, Journal of JSEE, Vol.54, No.5, pp.80–83, 2006.

[11] http://julius.osdn.jp/ (accessed on 16th March 2017)

[12] Tomas Mikolov, Kai Chen, Greg Corrado, Jeffrey Dean, Efficient Estimation of Word Representations in Vector Space, arXiv:1301.3781, 2013.

[13] Norman Alm, Richard Dye, Gary Gowans, Jim Campbell, Arlene Astell, Maggie Ellis, A Communication Support System for Older People with Dementia, Computer, Vol. 40, Issue 5, 2007.

[14] Masahiro Shiomi, Daisuke Sakamoto, Takayuki Kanda, Carlos Toshinori Ishi, Hiroshi Ishiguro, and Norihiro Hagita, A semi-autonomous communication robot - A field trial at a train station, 3rd International Conference on Human-Robot Interaction, 2008.

[15] Wataru Sunayama, Yuki Shibata, and Yoko Nishihara, Topic Recommendation Method related to a Present Topic for Continuing a Conversation, Information Engineering Express, Vol. 3, No. 1, 2017.

[16] Tian Xu, Hui Zhang, and Chen Yu, See You See Me: The Role of Eye Contact in Multimodal Human-Robot Interaction, ACM Transactions on. Interactive Intelligent Systems, Vol.6, No.1, pp.2:1–2:22, 2016.

[17] Ryosuke Yamanishi, Satoshi Miyake, Yoko Nishihara, Junichi Fukumoto, Visualization of Profile Affinity with Thesaurus and Web Contents for Communication Support, CHI 2016 Symposium on Emerging Japanese HCI Research Collection, 2016.

[18] Ye Kyaw Thu, Takuya Ishida, Naoto Iwahashi, Tomoaki Nakamura, and Takayuki Nagai, Symbol Grounding from Natural Conversation for Human-Robot Communication, Proceedings of the 5th International Conference on Human Agent Interaction, pp. 415–419, 2017.