RESEARCH NOTE

Analysis of Speech Dialogue to Detect Active Conversation and Lapse in Conversation toward Development of Conversation Support Robot after Co-occurrence and Mutual Assistance Matching

Reona Gomi, Hidekazu Aizawa, Eri Sato-Shimokawara and Toru Yamaguchi

Graduated school of System Design, Tokyo Metropolitan University

6-6 Asahigaoka, Hino-shi, Tokyo 191-0065, Japan

E-mail: {gomi-reona,aizawa-hidekazu}@ed.tmu.ac.jp

{eri,yamachan}@tmu.ac.jp

Abstract The formation of human relationships is important to form communities. As a supporting system to strengthen the bonding among community members, we firstly tried to connect users by co-occurrence and mutual assistance matching to form relationships. Then, we brought them together. At the time of this experiment, users can easily carry on animated conversations if a common topic is suggested to them. In this study, we utilized a questionnaire to acquire user's personal characteristics for co-occurrence and mutual assistance matching. Next, we conducted a conversation experiment via Skype to investigate activation and inactivation during user's conversations by using speech dialogue and the LF/HF ratio calculated from user's heart rates. During the conversation experiment, matched users made conversations for 10 min on each common topic based on their interests suggested. The result showed that users could easily have animated conversations when a common topic was suggested to them. In addition, we investigated acoustic features, the overlap of utterances, and the number of utterances, backchannel feedbacks, and laughter.

Keywords: conversation support, human relationship, co-occurrence and mutual assistance, active conversation, lapse in conversation

1. Introduction

In modern society, there are more opportunities for people to meet together owing to the widespread availability of the Internet and SNS. We should be able to build up and develop human relationships more easily

than before. On the other hand, there are many people isolating from their local communities, such as elderly people. Developed countries around the world faced serious problems of rapid aging and a low birthrate. In 2050, people over 60 years old will account for one in five of the world's population [1]. According to the

Fig. 1 An image about utilization of a trailer house

Ministry of International Affairs and Communications of Japan, the population of over 65 years old accounted for 27.3% of the total population of Japan in 2016 [2]. Moreover, elderly people who live alone and require nursing care have been continuously increasing. They often face the two major problems of isolations from their local societies and lack of communication with other people. Facing natural disasters, they may not be rescued because of their isolation. Involvements of such isolated elderly people into formation and vitalization of local communities is considered to be an effective approach for solving this isolation problem. For example, the trailer house shown in Figure 1 has been proposed. In daily life, the trailer house can be utilized as a kind of community houses for supporting formations of better communities, as well as to support the community in disasters. As mentioned before, it is important to form human relationships and better communities in people's daily lives. For all these reasons, it is easy to help each other when disasters occur.

There is a system which was developed by some Japanese researchers to prevent social isolation with self-efficacy scale of community-dwelling older people [3]. We can visualize monitoring networks on a neighborhood scale as part of support networks in community-scale. Such a system from being overlooked system prevents that matching based on isolated people.

In our consideration, 'co-occurrence' and 'mutual assistance' are beneficial for supporting community strength. In this research, we defined 'co-occurrence relationship' as a relationship connected by common hobbies and topics. Moreover, we define 'mutual assistance relationship' as a relationship in which two people with different strengths and weakness matching based on. For proceeding both 'co-occurrence' and 'mutual assistance', we need to collect personal characteristics, including each user's favorites and abilities. However, it can be difficult to support people who are isolated from their communities. We expect to

promote the mutual assistance in the process of community that includes those isolated people. To this end, human relationships are required. To help form relationships, it is important for people to be connected by common interests and to be able to help each other on the basis of their personal characteristics. This can be facilitated by a matching system. Moreover, we need a good support system to facilitate smooth conversation when people meet face to face.

Our final goal is to develop a conversation support robot such as that shown in Figure 2. Firstly, this system conducts 'co-occurrence' and 'mutual assistance matching' by using the acquired information of user's personal characteristics. Secondly, this system supports conversations between users connected by the matching. Thirdly, this system detects active conversation and a lapse in conversations in real time. If the system detects active status in the user's conversations, the support robot will start monitoring the user's conversations. On the other hand, if this system detects in a lapse activation in the user's conversation, the support robot will suggest suitable topics.

In conversations among users, a variety of non-verbal information is used for promoting smooth communications. We mainly focus on two items of non-verbal information to develop the conversation support robot. One is an acoustic feature such as volume control of speeches or the tone of voice. The other is biological information : the synchronization of user's heart rates. In this study, we conducted a co-occurrence and mutual assistance matching and conversation experiment via Skype. We analyzed user's dialogue speeches to detect active conversation and a lapse in conversation as described in detail in Section 3 to Section 4.3. Moreover, we analyzed the synchronization levels of user's heart rates during their conversations and investigated its relationship with active conversation and a lapse in conversation as shown in detail in Section 4.4.

Fig. 2 An image of the support system by the conversation support robot based on co-occurrence, mutual assistance matching and personal characteristics

2. Related Research

2.1 Matching between users

Recently, there have been much researches on support for forming human relationships. For example, there is a system in which a complex network theory is combined with the genetic algorithm of Pareto optimality to recommend friends [4]. Through the combination of algorithms and social networks, this system can give recommendations suited to individual needs. Moreover, there is a research on improving such recommendation systems using relationships in SNS (i.e. social networks) [5]. Also, there is a system which refers to the features of co-author networks for the formation of research groups [6]. Moreover, another system utilizes the common feature detected in pictures to form recommendations in addition to the information of genders and affiliations of LinkedIn accounts [7]. In this way, there are various existing studies on information analysis and recommendation systems based on SNSs.

However, we focused on a user's daily life in this research. Therefore, we needed to analyze a user's private information such as their hobbies, interests, and abilities. As one of the elements to connect people, we focused on their interests and abilities. For example, it is important to learn how user's qualifications were acquired and careers were chosen to estimate user's abilities. By controlling our robot's utterances, we were able to ask 'Do you like ~?' or 'Are you interested in ~?'. Therefore, we were able to focus exclusively on obtaining the necessary information. This robot conversation was effective for presenting recommendations as well as asking the users questions. There is a similar research on recommendation methods based on user's interests in SNSs [8].

The most conventional survey method of personal characteristics is a questionnaire. However, a questionnaire is difficult to conduct frequently. Moreover, it does not allow us to inspect transitions of user's interests. We investigated how to acquire their personal characteristics by using our questionnaire. We apply the results to the robot dialogue updates and the utterance control. In this research, we focused on conversation support after robot questioning. We investigated user's interests by the questionnaire and matched them as conversation mates according to the questionnaire results.

In our previous research, we developed a matching system for co-occurrence and mutual assistance [9]. We acquired each participant's personal characteristics such as abilities/disabilities and interests. According to the results, we selected user pairs to use this system. Moreover, in our previous research, we conducted a conversation experiment for users connected by mutual assistance matching with a help of mediators [10]. Topics

suggested by a mediator in their conversation topics are likely to support smoother communication. This result was shown by the physiological phenomenon: the synchronization of the heart rates.

2.2 Toward the implementation of conversation support system

After assigning pairs on the basis of co-occurrence and mutual assistance matching, opportunities to form co-occurrence and mutual assistance relationships beyond greetings and self-introductions are needed. Thus, we aim to develop a conversation robot which can act as a mediator in user's conversation. The robot is expected to provide proper suggestions of co-occurrence and/or mutual assistance topics in time the end of greetings and self-introductions by the users. This system leads active conversations between the users.

As research related to the conversation robot, there is a research on a robot that communicates naturally in accordance with different behaviors [11]. This robot is capable of several speaking patterns such as a daily dialogue and information assistance. There is also a survey on what robot characteristics are suitable for communicating with humans [12]. Recently, a method of calculating a robot's emotional empathy models on the basis of human emotions has been proceeded [13]. Such research is aiming to achieve natural communication between human and a robot. At this point, these researches are somehow similar to our research. However, we propose that our conversation robot acquire a user's personal characteristics and match with their daily conversations at the same time. This is the unique point of our research.

There is a research on a conversation robot supporting people who have seldom chances to communicate with others such as elderly people who live alone [14]. In the other case, a robot communicates with users who are watching TV, and encourages them to start communication. In the same way, there is also a conversation robot to share past experiences with people afflicted with dementia [15]. In this way, there are many researches on conversation robots as conversational partners for humans. Conclusively, we aim to acquire personal characteristics through user's conversations with our robot system. In this situation, the robot can be a conversational partner in the phase of acquiring personal characteristics, similar to the related research. However, it is not enough to merely make converse as a conversational partner. The robot is unable to judge whether the users are interested in suggested topics or not, although the robot has to understand whether users are interested in the topic of the conversation. In particular, to support human-human conversation, a robot must understand whether users are interested in the topic. Therefore, we conducted an experiment and analysis to

search for a barometer which is useful in supporting conversations from various acoustic features by observing human-to-human conversation. For a robot system that supports conversations, there has been a research focused on the user's acoustic features. Moriya et al. proposed a method for automatically displaying an avatar of non-verbal information such as glances and facial expressions [16]. Their research was aimed at linking humans and robot. We aim for robot systems that can mediate between users, so our research is different from theirs.

3. Survey of Personal Characteristics and Experiment to Acquire Conversation Data

In this work, firstly, we used a questionnaire to acquire personal characteristics of users. Secondly, we conducted both co-occurrence and mutual assistance matching using the acquired data of their personal characteristics and selected user pairs for a conversation experiment via Skype. In the Skype conversation experiment, we collected dialogue speech data for each user by recording the user's conversations. Students in our laboratory were recruited as examinees for these experiments. In fact, it is difficult to gain cooperation from elderly people as examinees, even though we pointed that elderly people were the good example in Section 1. This is because one attendee converse with multiple people on the matching results. However, conversation topics may differ among age groups. Therefore, we considered a survey of suitable topics for each age group after developing youths-focused of conversation system.

3.1 Survey of personal characteristics

We used a questionnaire to acquire user's personal characteristics. In our previous research, we adopted two methods including the questionnaire and conversations with a robot [17]. There were no differences between the results of these two methods. In this research, our primary aim is making the robot be able to support human-to-human conversations by digitalizing the observed users' conditions during conversation, so we adopted the questionnaire. We prepared 60 questions comprising 30 questions on user's abilities and another

30 questions for user's interests. Table 1 shows the topics in the questionnaire. In our research, ability means the user's strong points. Interest means the level for which the user is interested in each topic. They answered the questions on a 5-point scale. We referred to 'Yahoo! Answers' for the categories and personal profiles of Facebook for interest category. The attendees in this experiment were 12 students in their twenties who belong to the same laboratory in our university.

3.2 Co-occurrence and mutual assistance matching

We utilized the matching method from our previous research [18]. First, a user was selected and was compared with other participants in terms of the ability. Then, levels of similarity were calculated by using all the ability lists other than the selected users. The levels of similarity were calculated by using the Pearson correlation. Table 2 shows the results of matching.

The users in a mutual assistance relationship help each other to overcome their respective weak points. For example, firstly, user A is good at cooking, while user B is not good at cooking, but is interested in cooking. Second, user B is good at DIY (i.e. do-it-yourself), while user A is not good at DIY, but is interested in DIY. In this case, user A can teach cooking to user B and user B can teach DIY to user A. Therefore, user A and user B are able to help each other with respect to their weak points. The mutual assistance matching system performs the results of co-occurrence matching and selects a suitably matched. The system refers to the others as those who have a low co-occurrence levels, i.e. whose similarity degree is less than 0.05, relative to the selected user. The system also refers to the user's interests. For example, one user may be good at sports. While another may be interested in sports but is not athletic. In this study, we selected other pairs to compare with co-occurrence or mutual assistance pairs. The other pairs gave responses of '3' for levels of interest in certain topics.

3.3 Conversation experiment via Skype - 10 minutes separated conversations

Using the results of a conversation experiment with natural language dialogue [19], we changed the timing of suggesting topics because it was necessary to eliminate variability in the time and the number of utterances for each topic of conversation. We conducted the

Table 1 Topics of the questionnaire to acquire personal characteristics

cooking	cleaning	Laundry	rising early	sewing
drive a car	tidying	conversation	drinking	ironing
reading books	karaoke	detailed work	taking pictures	games
instrument	programming	DIY	Shopping	outdoors
teaching	sports	diary	Planning	joining events
acquisition of qualification	language learning	Fitness	drawing	taking care of pets

conversation experiment for seven pairs. Figure 3 shows a view of the experiment. Three members attended the experiment including of two users conversing and a mediator who instead the support robot.

The users were each situated in a quiet room. Users paired by the matching had conversations via Skype on a PC. Then, the conversation was recorded via a voice recorder placed next to the PC during this conversation experiment at a constant distance from the user. A mediator was able to hear the user's conversations. There were three co-occurrence pairs, two mutual assistance pairs, and two other pairs. A mediator suggested 3 topics to each pair by text chat and allotted 10 minutes for each topic which are listed in Table 3. We prepared 2 topics of co-occurrence and mutual assistance and 1 another topic selected randomly among those unrelated to co-occurrence and mutual assistance. The topics of the other pairs were selected at random among all the topics.

4. Results and Analysis of Dialogue Speech Data

It is important to consider the appropriate time of changing the conversation topics. Therefore, we tried to classify only the user's utterance intervals by topic. However, we found it difficult to estimate dialogue contents from dialogue speech data from the previous analysis results [19]. This time, we prepared dialogue speech data because we need to acquire acoustic features of all dialogue speech data. We tried to classify dialogue speech data into three stages: beginning of conversation, active conversation and a lapse in conversation.

4.1 Preparation—splitting every 5 seconds

We equally splitted all the dialogue speech data every 5 seconds. This is because we must analyze the dialogue speech data periodically in real time when we utilize the conversation support robot. Using the Praat software [20]

we acquired 8 acoustic features: acoustic pressure effective values, maximum and minimum acoustic pressures, a range of acoustic pressures, pitches in average, maximum and minimum pitches, and a range of pitches.

Fig. 3 A view of the conversation experiment

Table 3 Example of topics suggested for each pair

	co-occurrence	mutual assistance	others
user pair	users 2,11	users 9,11	users 7,11
topic 1	cooking	<u>fitness</u>	joining events
topic 2	<u>reading books</u>	<u>license</u>	sports
topic 3	<u>karaoke</u>	sewing	taking care of pets

Note: The double underlines = Topics of both co-occurrence and mutual assistance

4.2 Classification by active and lapse in conversation

We classified dialogue speeches by active/lapse in conversation to investigate whether we can utilize acoustic features to judge active/lapse for conversation support. We verified the detection of the start of conversation, active conversation, and lapse in conversation. Compared with the case of natural conversation [19], we predicted to ensure a sufficient

Table 2 Matching results

	user2	user3	user4	user5	user6	user7	user8	user9	user10	user11	user12
user1	0.428	0.172	0.186	0.185	0.372	0.063	-0.136	0.044	0.047	0.201	-0.077
user2		0.218	0.179	0.000	0.060	0.180	0.299	0.115	0.265	0.504	0.230
user3			0.271	0.312	-0.136	-0.030	0.272	0.009	-0.019	0.382	0.475
user4				0.488	-0.379	-0.438	0.238	0.191	0.193	0.069	0.349
user5					-0.131	-0.296	0.085	0.040	0.080	0.037	0.172
user6						0.362	-0.400	0.018	0.004	0.013	-0.215
user7							-0.091	-0.152	0.180	0.370	-0.390
user8								0.308	0.054	0.401	0.158
user9									0.118	0.124	0.030
user10										0.126	0.391
user11											0.264

Table 4 Sample transcript of active conversation

user	utterance
B	4 時むっちゃ明るいもん, [朝 4 時]
A	(((laughing)))に寝る]んでしょ?
AB	((laughing A and B))
A	[もうきれそう]になるよね朝 4 時((laughing))
B	[そうそうそう]((laughing))
A	[今から寝るん]だけどもみたいな((laughing))
B	[そうそうそう]((laughing))
A	なんで明るく[なってるんだよ]((laughing))
B	[そうそうそう][今からさ]
A	[ふざけんなよ]みたいな((laughing))

Note: “[=” start of overlap, “]” = end of overlap, “(())” = description of circumstance

Table 5 Sample transcript of lapse in conversation

user	utterance
B	すごく昔のことですね
A	そうですね, だいぶ前でしたもんね映画自体が
B	はい, そうですね
A	((silence for 1 sec))Bさんは最近何見ました? ジュラシックワールドですか?
B	ジュラシックも最近見ますけど, この前は, でも 最近忙しいですね.
A	あー
B	このあと 8 月は外国の映画をたくさん見に行き たいです
A	あー
B	((silence for 0.8 sec))じゃあミッションインポッシ ブルとか
A	ああはいそれも見たいです

Note: “[=” start of overlap, “]” = end of overlap, “(())” = description of circumstance

number of utterances. We manually attached three tags to dialogue speech data by listening to the conversation data. The three tags mean “introduction”, “active”, and “lapse” respectively. First, we defined the intervals of the start of each topic as “introduction”. Second, we defined the intervals including several instances of overlap and talking with laughing as “active”. Table 4 shows a sample transcript of active conversation. Third, we defined the intervals including several instances of silence before utterance and no overlap and no laughing as “lapse in conversation”. Table 5 shows a sample transcript of lapse in conversation.

We acquired acoustic features using Praat. Next, we classified the acoustic features data for each user by 10-fold cross-validation utilizing the multilayer perceptron in WEKA software [21]. WEKA showed high accuracy at the same classification problem utilizing the acoustic

Table 6 Classification result by active conversation and lapse in conversation.

co-occurrence	1	user12	58.1%
		user3	51.8%
	2	user2	54.1%
		user11	60.2%
	3	user1	75.9%
		user2	68.2%
mutual assistance	1	user11	68.6%
		user9	62.5%
	2	user7	49.7%
		user4	68.9%
other	1	user2	59.2%
		user7	59.2%
	2	user11	62.7%
		user7	67.0%

feature in the previous research [22]. Table 6 shows the classification accuracy rate for the acoustic feature we tagged manually. The highest accuracy rate is 75.9% from user1. All of these accuracy rates are not high, so we conclude that the use of only the acoustic feature is unsatisfactory. Next, we investigated the time taken for a conversation to become active. The results are shown in Figure 4 for each topic, user, and conversation time. These results were obtained by calculation utilizing manually tagged data. The blue zone in Figure 4 shows the interval of the beginning of conversation. The orange zone in Figure 4 shows the interval of active conversation. ‘t1’, ‘t2’, and ‘t3’ referring to each topic number. We focus on the three pairs in the red frame in Figure 4. All of these pairs include user11. We investigated whether there are differences between the user’s relationship and active/lapse or not. The users 9 and 11 pair were suggested an other topic after two topics related to mutual assistance. Moreover, the users 2 and 11 pair were suggested two topics related to co-occurrence after the other random topic. We focus on the time users need to reach activation in topic3 for all pair in red frames. The time to reach active conversation in topic3, which of co-occurrence and mutual assistance pairs are shorter than that of the other two pairs. Therefore, the time that users need to active conversation is likely to become shorter by co-occurrence and mutual assistance relationships.

4.3 Verification of turn-taking, overlap and active conversation interval

We investigated overlap and turn-taking in a conversation for the three pairs indicated in Figure 4.

Fig.4 Correlation graph between the numbers of tags rates for each topic and conversation time

Table 7 Number of tags for each pair

	user ID	back-channel feedback (times)	laughing (times)	utterance (times)
co-occurrence	user2	243	53	183
	user11	128	49	243
mutual assistance	user9	110	22	177
	user11	81	74	170
the other	user7	169	90	221
	user11	139	58	234

Table 8 Number of overlaps for each topic and each pair

co-occurrence users 2,11	topic	cooking	<u>reading books</u>	<u>karaoke</u>
	times	127	127	124
mutual assistance users 9,11	topic	<u>fitness</u>	<u>license</u>	sewing
	times	87	85	97
the other users 7,11	topic	joining events	sports	taking care of pets
	times	120	117	126

Note: double-underline = topics of co-occurrence and mutual assistance

Overlap usually occurs when the dialogist is interested in conversation [16]. Moreover, we can judge a conversation to be animated when many turns are taken [23]. In particular, we estimated we were able to know the trend of listener and speaker from the number of utterances and back-channel feedbacks. Therefore, we adopted three new tags: back-channel feedback, laughing,

Table 9 Durations of active and lapse periods for each topic and each pair

pair	duration (s)	tag	topic
co-occurrence users 2,11	7.830	active	cooking
	6.903	active	<u>karaoke</u>
	14.634	active	
mutual assistance users 9,11	58.881	lapse	<u>fitness</u>
	53.686	active	
	24.610	lapse	<u>license</u>
	30.198	active	
	19.374	active	sewing
the other users 7,11	6.894	lapse	joining events
	13.334	active	sports
	12.771	active	taking care of pets
	23.788	active	

Note: double-underline = topics of co-occurrence and mutual assistance

and utterance. Table 7 shows the number of tags for each pair. In the co-occurrence and the other pair, we focus on the number of back-channel feedback tags and utterance tags. In both pairs, the number of utterance tag of user11 is larger than that of the pair mate. Moreover, in both pairs, the number of back-channel feedback tags of the pair mate is larger than that of user11. Therefore, we estimate that user11 tends to act as a speaker in co-occurrence and the other pair.

Next, we investigated the time of overlaps. When a user's utterance overlaps another's utterance or back-channels feedback, we defined it as an overlap. Table 8

shows the number of overlaps for each topic and each pair. We focus on the mutual assistance pair and the other pair. In both pairs, the number of overlaps for topic3 is the largest. In particular, the overlap result shows the tendency of which increment for the mutual assistance pair is larger than that for the other pair. Therefore, there is possible co-occurrence and mutual assistance relationships are considered to be valuable for supporting formation of human relationships.

We manually tagged active and lapse by listening to dialogue speech between two persons. We defined the durations of period of both evaluations as final periods of active and lapse. Table 9 shows active and lapse duration of periods. These periods are calculated as the start time of a period subtracted from the finish time. We focus on the co-occurrence pair. For each successive topic, the active period becomes longer. The active periods for the third topic of co-occurrence are the longest. Therefore, conversation easily becomes active for topics on users' common interests.

4.4 Heart rate analysis in the durations of active periods

The active phenomenon is not only overlap. When conversation is animated, physiological phenomena such as HRV (heart rate variability) and breathing interval synchronize between participants. Such synchronized biorhythm is called 'entrainment'. In the previous section, we focused on an acoustic feature in non-verbal information. In this section, we focus on biorhythm. The user wore a heart rate sensor on the ear to monitor the heart rate during conversation. We analyzed the shift in the heart rate interval being an indicator of stress, when a topic was suggested during the conversation by a mediator. In an electrocardiogram, the sharpest peak is called the R-wave. The interval between two consecutive R-waves is called the RR interval. In this study, RR interval data were collected at a sampling rate of 1Hz. The power spectral density is calculated from the RR

Fig. 5 Variability of the LF/HF ratio before interpolation (top) and after interpolation (bottom)

Fig. 6 The LF/HF ratio graph for users9 and 11 pair

interval data with 100 s window size and shift size of 1 s. It has been confirmed that there are two peaks in a power spectrum of HRV wave [24]. The low-frequency band (0.05 – 0.14Hz) is called LF. This region reflects blood-pressure fluctuation. Moreover, LF is associated with both the sympathetic and parasympathetic nerves. The high-frequency band (0.15 – 0.40Hz) is called HF. This region reflects the breath-interval variability. In addition, HF is associated with only the parasympathetic nerve. Therefore, the LF/HF ratio is utilized as an assessment of sympathetic nerve stress [25]. There is a study in which the LF/HF ratio was calculated by real-time processing [26]. Therefore, a robot may be able to detect the user's mental state from the LF/HF ratio depending on the presented topic. In this study, we compared the difference in the LF/HF ratio in the presented topic, and analyzed the variability of the LF/HF ratio as well.

To counter noise, a linear interpolation is performed for error of the heart rate value. Figure 5 shows a graph of the LF/HF ratio. A normal R-R interval is determined from the normal heart rate interval. In rare cases, out-of-order values were outputted as R-R interval because of facial movement during conversation. Heart rates under 40 and over 120 are defined as sickly bradycardia and tachycardia respectively [27]. We converted to heart rate interval and defined the normal interval to be from 500 to 1,333ms. For an example of an out-of-order value, we measured 252ms that might be affected by facial movement. Moreover, the major methods of interpolation to determine the LF/HF are linear interpolation and spline interpolation [28]. In this research, we utilized linear interpolation because it was necessary to utilize a simple method for real-time processing.

The top of Figure 5 shows a wave before interpolation. The bottom shows a wave after interpolation. At the part with a red background, there is a difference between the waves before and after interpolation. In the waves before interpolation, because of noise, the LF/HF ratio is lower than those in other regions. On the other hand, in the wave after interpolation, there is no difference among the other regions.

As an example, Figure 6 shows the LF/HF ratio graph for the users 9 and 11 pair. In the region with the red background in Figure 6, we can confirm that the

Table 10 Average LF/HF ratio of for each topic for co-occurrence pair

pair	topic	first-shown user's LF/HF average	second-shown user's LF/HF average
user3 user12	programming(c.o.)	2.593	1.819
	waking up early(c.o.)	2.452	2.259
	conversation(other)	1.830	1.652
user1 user2	cleaning(c.o.)	-	-
	drawing(other)	-	-
	game(c.o.)	2.267	1.047
user2 user11	cooking(other)	1.612	1.753
	reading books(c.o.)	1.258	1.699
	karaoke(c.o.)	1.735	2.060

Table 11 Average LF/HF ratio of for each topic for mutual assistance pair

pair	topic	first-shown user's LF/HF average	second-shown user's LF/HF average
user9 user11	fitness(m.a.)	1.831	3.758
	license(m.a.)	1.329	3.324
	sewing(other)	1.968	2.339
user4 user7	reading books(m.a.)	2.915	2.023
	license(other)	2.258	1.692
	drawing(m.a.)	1.970	1.605
user1 user10	detailed work(other)	2.652	2.195
	programming(m.a.)	2.950	1.127
	waking up early(m.a.)	3.012	1.663

Table 12 Average LF/HF ratio of for each topic for the other pair

pair	topic	first-shown user's LF/HF average	second-shown user's LF/HF average
user3 user7	DIY(other)	2.386	1.897
	tidy(other)	2.721	1.550
	instrument(other)	1.825	1.491
user2 user7	diary(other)	2.244	1.626
	sewing(other)	1.712	1.637
	shopping(other)	1.702	1.497
user7 user11	joining events(other)	1.776	1.833
	sports(other)	2.255	1.767
	taking care(other)	2.107	1.340

transition of heart rate is synchronized. During this interval, the topic “sewing” was suggested as the other topic after the two topics of mutual assistance. In the questionnaire, the users indicated their depth of interest to be 3. However, one user is good at sewing while the other is not. We found that the topic “sewing” was nearly equal to a mutual assistance topic. Table 10 to 12 shows

the average LF/HF ratios for co-occurrence, mutual assistance, and the other pair respectively. In Table 10, “c.o.” indicates a co-occurrence topic, and “-” means missing data. In Table 11, “m.a.” indicates a mutual assistance topic.

Fig. 7 Overlap time graph (top) and LF/HF ratio correlation graph (bottom) for the users 2 and 11 pair

Fig. 8 Overlap time graph (top) and LF/HF ratio correlation graph (bottom) for the users 9 and 11 pair

Moreover, Figure 7 to 9 shows the LF/HF ratio correlation graph and overlap time graph for each pair. The time at which we were able to measure the LF/HF was over 100 s after starting the experiment. At the top of Figure 7 to 9, the top graph shows overlap time every 100 s and the bottom graph shows the LF/HF correlation. The horizontal axis of those show the intervals as 101—200, 201—300, 1701—1800. The overlap interval means the number of seconds that the utterances of two persons was overlapped. Figure 7 to 9 shows that overlaps coincide with the LF/HF ratio several times, as marked with red circles. This overlap phenomenon is often seen when users are interested in the topic of the conversation. The correlation between the overlap occurrence rate and the LF/HF ratio was found to be 0.4 using the CORREL function in Excel, which indicates a weak correlation. We investigated the relationship between the overlap time

Fig. 9 Overlap time graph (top) and LF/HF ratio correlation graph (bottom) for the users 7 and 11 pair

interval and LF/HF correlation. During active conversation or lapse in conversation, two phenomena may occur: positive/negative perception of users toward conversation and the effects on biological signals such as heart rates. However, there is a time lag when multiple biological reactions appear [29]. In this research, we estimate there is a time lag between positive/negative aesthesis of users toward conversation and effects to biological signal as heart rate. However, we estimate that the correlation coefficient is improved by considering the time lag. Therefore, we must investigate the time lag to determine whether it depends on the users and topics.

The average of all participants' LF/HF ratio is 2.182 for topic1, 1.999 for topic2, and 1.862 for topic3. Regardless of the topic contents, sympathetic nervous activity tended to decrease as the experiment progressed ($p < 0.02$ between topic1 and topic3, calculated with a Wilcoxon signed-rank test). Prior to experiments, the user might be nervous. Besides, the sympathetic nervous system exerts more effect than the parasympathetic nervous system. As the experiment progresses, the users could relax, would explain the observed trend. Furthermore, the same trend was also observed in the result of the questionnaire on ease of talking Table 14. In this experiment, ten participants responded to the questionnaire. According to the result, there is no respondent that topic3 is the worst conversable. Rather, all respondents said that indicated topic3 was the most conversable. Thus, when considering the order of topics, it is better to present the topic that garners the most interest last.

Table 13 Correlation between overlap occurrence rate and the correlation graph of the LF/HF ratio

relation	user ID	correlation
co-occurrence	user2,11	0.05
mutual assistance	user9,11	0.40
the other	user7,11	-0.53

Table 14 The result of the questionnaire on ease of talking

topic	worst ease of talk	most ease of talk
topic1	7	0
topic2	3	0
topic3	0	10

5. Summary

Forming stronger human relationships is vital for better community formation. As a supporting system to strengthen bond among community members, we firstly tried to connect users by the co-occurrence and mutual assistance matching to form relationships. Then, we bring users together, and at which time, users may easily fall into an animated conversation upon the suggestion of a common topic. We conducted a questionnaire to acquire user's personal characteristics, and co-occurrence and mutual assistance matching. Next, we conducted conversation experiments via Skype to investigate active conversation and lapses in conversation by using speech dialogue and the LF/HF ratio calculated from the user's heart rates. During the experiment, matched users had conversations centering on a common topic for 10 min. Users easily fell into an animated conversation when a common topic was suggested. We also investigated acoustic features, overlap of utterances, and type of speeches from the speech dialogue, which includes utterances, backchannel-feedbacks, and laughing.

We found that it is likely to detect active conversation and lapse in conversation on the basis of information of the heart rate and speech dialogue data. We investigated the types of utterance (back-channel feedback, utterance, and laughing), overlap time, and the durations of active conversation periods. We focused on the number of utterances and back-channel feedbacks. In the co-occurrence and the other pair, we focus on the number of back-channel feedback tags and utterance tags. In both pairs, the number of utterance tag of user11 is larger than that of the pair mate. Moreover, in both pairs, the number

of back-channel feedback tags of the pair mate is larger than that of user11. Therefore, we assume that users have the role of speaker or listener. The increments of overlap time for each topic for the mutual assistance pair were larger than that for the other pair. Moreover, the increment of the active conversation interval for each topic for the co-occurrence pair was larger than that for the other pair. Therefore, it is considered that co-occurrence and mutual assistance relationships may contribute to active conversation. Then, we investigated the LF/HF ratio calculated from heart rate during conversation for the mutual assistance pair. It was confirmed that synchronized LF/HF transition was observed in the interval when the other topic "sewing" was suggested. This topic was suggested after two topics of mutual assistance. To confirm user's answers of the questionnaire, we found users were connected by this topic as nearly mutual assistance relation.

We found that as the experiment progressed, the users become relaxed. The same trend was also observed in the results of the questionnaire about the ease of talking. All respondents had answered that topic3 was the easiest one to talk. From this, when thinking about the order of topics, it is better to present the topic that includes the most active conversation last. Moreover, there is likely to be some connection between active conversation and lapse in conversation, and co-occurrence and mutual assistance relationships. This is because the transition of the LF/HF ratio is similar to the transition of the overlap time interval in co-occurrence and mutual assistance pairs. For example, a high LF/HF ratio means the occurrence of the entrainment phenomenon, which indicates that the users are interested in conversation and its perpetuation. Moreover, a long overlap means frequent and the conversation was animated. The detection accuracy of active conversation and lapse in conversation is not very high. We assume two causes of this result. First, the topics we prepared for the experiments are likely difficult to talk about. So, we will have to review the questionnaire. Second, in some situations that users exhibit a strong trend of speaker and listener, active conversation may be difficult. Therefore, we need to reselect topics for experiments and consider the intensity level of the trend of speaker and listener.

In the future, to judge active conversation and lapse in conversation in real time, we detected both dialogue speech and heart rate. The use of these two factors was considered to ensure effective conversation support. In addition, in the future, as our further work, we will investigate whether the order effect of topics has an effect on active conversation and lapse in conversation. Therefore, a robot intervenes in a user's conversation at a suitable timing will become possible.

References

- [1] <http://www.who.int/ageing/events/international-day-older-persons/2017/en/> [accessed Jan. 25, 2018]
- [2] <http://www.stat.go.jp/english/data/handbook/c0117.htm> [accessed Jan. 25, 2018]
- [3] E. Tadaka, A. Kono, E. Ito, Y. Kanaya, Y. Dai, Y. Imamatsu and W. Itoi: Development of a community's self-efficacy scale for preventing social isolation among community-dwelling older people (Mimamori Scale) , BMC Public Health, Vol.16, No.1, pp.1-9, November 2016.
- [4] J. Naruchitparames, M.H. Gunes and S.J. Louis: Friend recommendations in social networks using genetic algorithms and network topology, IEEE Congress of Evolutionary Computation (CEC 2011), pp.2207-2214, June 2011.
- [5] H. Tian and P. Liang: Improved recommendations based on trust relationships in social networks, Future Internet, Vol.9, No.1, Article Number 9, March 2017.
- [6] Q. Yu, C. Long, Y. Lv, H. Shao, P. He and Z. Duan: Predicting co-author relationship in medical co-authorship networks, PloS One, Vol.9, No.7, e101214, July 2014.
- [7] S. Tifferet and I. Vilnai-Yavetz: Self-presentation in LinkedIn portraits common features, gender, and occupational differences, Computers in Human Behavior, Vol. 80, pp.33-48, March 2018.
- [8] R. Logesh and V. Subramaniaswamy: A reliable point of interest recommendation based on trust relevancy between users, Wireless Personal Communications, Vol. 97, No.2, pp.2751-2780, November 2017.
- [9] R. Gomi, A. Suzuki, E. Sato-Shimokawara and T. Yamaguchi: Analysis of dialogue for acquiring personal characteristics toward co-occurrence matching, 2015 Conference on Technologies and Applications of Artificial Intelligence (TAAI 2015), Taiwan, pp.206-212, November 2015.
- [10] R. Gomi, H. Aizawa, E. Sato-Shimokawara and T. Yamaguchi: Analysis of dialogue speech between users tied by co-occurrence and mutual assistance matching, Proceedings of the 2017 JSME Conference on Robotics and Mechatronics (ROBO-MECH2017), CD-ROM No. 2A1-L09, May 2017 (In Japanese).
- [11] T. Obo, T. Takeda, J. Botzheim and N. Kubota: Human-friendly communication for smart device interlocked robot partners, IFAC-PapersOnLine, Vol. 49, No.19, pp.132-137, December 2016.
- [12] A. Abe and M. Hayashi: On communication assistance via bots – towards IMDJ, 20th International Conference on Knowledge Based and Intelligent Information and Engineering Systems (KES 2016), Procedia Computer Science, Vol. 96, pp.1657-1665, September 2016.
- [13] J. Woo, J. Botzheim and N. Kubota: Emotional empathy model for robot partners using recurrent spiking neural network model with Hebbian-LMS learning, Malaysian Journal of Computer Science, Vol. 30, No.4, pp.258-285, December 2017.
- [14] M.M.M. Peeters and M.A. Neerinx: Human-agent experience sharing: Creating social agents for elderly people with dementia, 24th ACM Conference on User Modeling, Adaptation and Personalisation (UMAP 2016), CEUR Workshop Proceedings, Vol. 1618, July 2016.
- [15] H. Minami, H. Kawanami, M. Kanbara and N. Hagita: Chat robot coupling machine responses and social media comments for continuous conversation, IEEE International Conference on Multimedia and Expo Workshop (ICMEW 2016), Category No. CFP16IEW-ART, July 2016.
- [16] Y. Moriya, T. Tanaka, T. Miyajima and K. Fujita: Study of activation level estimation using phonetic information during voice chat, Journal of Human Interface Society: Human Interface, Vol.14, No.3, pp. 283-292, 2012 (in Japanese).
- [17] R. Gomi, T. Kaneko, A. Suzuki, E. Sato-Shimokawara and T. Yamaguchi: An analysis of human-robot conversation for acquiring personal characteristics toward mutual assistance matching, 2015 RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing (NCSP'15), Kuala Lumpur, Malaysia, 28PM1-2-1, February, 2015.
- [18] H. Aizawa, S. Iwasaki, R. Gomi, E. Sato-Shimokawara and T. Yamaguchi: Heart rate analysis in a conversation on video chat for development of a chat robot supporting to build a relationship, 2017 IEEE/SICE International Symposium on System Integration (SII 2017), Taiwan, WeC4.8, December 2017.
- [19] R. Gomi, H. Aizawa, E. Sato-Shimokawara and T. Yamaguchi: An analysis of dialogue based on co-occurrence and mutual assistance to develop conversation support robot, Proceedings of the 5th International Workshop on Advanced Computational Intelligence and Intelligent Informatics (IWACIII2017), Beijing, November 2017.
- [20] P. Boersma and D. Weenink. Praat: Doing phonetics by computer [Computer program]. Version 6.0.37,

retrieved 3 February 2018 from <http://www.praat.org/> [accessed Dec. 20, 2017].

- [21] E. Frank, M. A. Hall and I. H. Witten: The WEKA workbench. Online appendix for data mining: Practical Machine Learning Tools and Technics, 4th edition, 2016.
- [22] C. Ikuta, E. Sato-Shimokawara and T. Yamaguchi: Analysis of dialogue speech to detect keyword toward acquiring topics for conversation robot, The 16th SICE System Integration Division Annual Conference (SI2015), pp. 1235-1238, CD-ROM No. 2C2-5, December 2015 (In Japanese).
- [23] C. De Vos, F. Torreira and S.C. Levinson: Turn-timing in signed conversations: Coordinating stroke-to-stroke turn boundaries, *Frontiers in Psychology*, Vol. 6, Article No.268, March 2015.
- [24] B.A. McSaykrs,: Analysis of heart rate variability, *Ergonomics*, Vol. 16, No.1, pp.17-32, January 1973.
- [25] T.A. Mellman, B.R. Knorr, W.R. Pigeon, J. Leiter and M. Akay: Heart rate variability during sleep and the early development of posttraumatic stress disorder, *Biological Psychiatry*, Vol. 55, No.9, pp.953-956, May 2004.
- [26] T. Iwamoto and S. Masuko: Lovable couch: Mitigating distrustful feelings for couples by visualizing excitation, *ACM International Conference Proceeding Series*, Vol.11, pp.157-158, March 2015.
- [27] <http://www.ncvc.go.jp/cvinfo/disease/arrhythmia.html> [accessed Aug.22, 2018]
- [28] https://www.jstage.jst.go.jp/article/fss/31/0/31_29/_pdf/-char/en [accessed Aug.22, 2018]
- [29] R. Nagasawa, H. Shimazu, K. Misawa and M. Yamashita: Cross-correlation coefficient with considering lag time between extracted RSA component of HRV and respiratory wave – A suggestion for exciting mode index, *Japanese Society for Medical and Biological Engineering*. Vol.48, No.2, pp. 181-188, April 2010 (in Japanese).

Reona Gomi received her B.S. degree in physics at 2013 from Hirosaki University, Aomori, Japan, and her M.S. degree in computer and information sciences in 2016 from Tokyo Metropolitan University, Tokyo, Japan. She is in a doctoral course in 3rd grade at Tokyo Metropolitan University of Information and Communications Systems Engineering.

Hidekazu Aizawa received the B.S. degree in computer and information sciences in 2017 from Tokyo Metropolitan University, Tokyo, Japan. He is in a master course in 2nd grade at Tokyo Metropolitan University of Information and Communications Systems Engineering.

Eri Sato-Shimokawara received her B.E., M.E., and D.E. in Systems Engineering Science from Tokyo Metropolitan Institute of Technology in 2002, 2004, and 2007. She was a Research Fellow of the Japan Society for the Promotion of Science (JSPS) from 2004 to 2007. She has been an Assistant Professor in the Faculty

of System Design of Tokyo Metropolitan University, Japan, since 2007. Her current research interests include human-machine interactions, multimodal interactions, soft computing, and intelligent robotics. She is a member of the Institute of Electrical and Electronics Engineers (IEEE), the Institute of Electronics Information and Communication Engineers (IEICE), the Japan Society for Fuzzy Theory and Intelligent Information (SOFT), and the Japanese Society for Artificial Intelligence (JSAI).

Toru Yamaguchi received his B.S. and M.E. in electrical engineering from Chiba University, Japan, in 1979 and 1981. He received his Ph.D. in computer science from Chiba University in 1992. He was with the System Software Laboratory of Toshiba Corp. in Tokyo, Japan, from 1981 to 1993 and worked in

the field of intelligent systems using fuzzy and neural network systems. He was also with the Laboratory for International Fuzzy Engineering (LIFE) Research as a chief researcher of the Fuzzy Associative Memory (FAM) project. Further, he was an Associate Professor at Utsunomiya University in Japan, from 1993 to 2000. He has been a full Professor at the Faculty of System Design of Tokyo Metropolitan University since 2000. He has also been participating in Precursorv Research

or Embryonic Science and Technology (PRESTO) of the Japan Science and Technology (JST) Corporation. He was an Associate Editor of IEEE Transactions in Industrial Electronics from 2000 to 2004, a Researcher of Ubiquitous Robot Technology Research center from 2004 to 2008, a Researcher of Institute of Industrial Science, The University of Tokyo from 2005 to 2009. He is Director of Community-centric System Research Center, Tokyo Metropolitan University, since 2015. His current interests are in soft computing, neural computing, fuzzy computing, and chaotic computing. Moreover, Dr. Yamaguchi was a recipient of the Technological Contribution Award of Society of Automotive Engineers of Japan, Inc. in 2011. He is the fellow in the Robotics Society of Japan. He is a member of the Institute of Electrical Engineers of Japan (IEEJ), the Japan Neural Network Society (JNNS), SOFT, Society of Instrument and Control Engineers (SICE), and the Robotics Society of Japan (RSJ).

(Received June 6, 2018; revised September 17, 2018)