Dialogue System Characterization by Back-channelling Patterns Extracted from Dialogue Corpus

Masashi Inoue, Hiroshi Ueno
Yamagata University
3-16, 4 Jyonan, Yonezawa
mi@yz.yamagata-u.ac.jp, tmk56575@st.yamagata-u.ac.jp

Abstract

In this study, we describe the use of back-channelling patterns extracted from a dialogue corpus as a mean to characterising text-based dialogue systems. Our goal was to provide system users with the feeling that they are interacting with distinct individuals rather than artificially created characters. An analysis of the corpus revealed that substantial difference exists among speakers regarding the usage patterns of back-channelling. The patterns consist of back-channelling frequency, types, and expressions. They were used for system characterization. Implemented system characters were tested by asking users of the dialogue system to identify the source speakers in the corpus. Experimental results suggest that possibility of using back-channelling patterns alone to characterize the dialogue system in some cases even among the same age and gender groups.

Keywords: keyword A, keyword B, keyword C

1. Introduction

Text-based communication in conversational formats has become increasingly common among people because of the widespread use of messaging applications. Accordingly, a growing need exists for text-based dialogue systems that can conduct natural conversation beyond single command utterances. Dialogue systems may someday become universal, which mean people will have to interact with different systems in different situations and for different purposes. It will then be natural for dialogue systems to employ different characters so that users can distinguish systems and enjoy different types of interaction.

Many studies have been conducted on generating characters through vocalization (e.g., (Fujie et al., 2004).) Compared to characterization in which personal traits are added to spoken sound, relatively few studies have been conducted adding linguistic personal traits such as vocabulary and style. The addition of linguistic characters to dialogue systems is often based on tailored rules (Nass et al., 1995) instead of statistical information obtained from corpora. As an example of statistical dialogue system characterization, one study has adapted the Big Five personality model to verbal expressions (Mairesse and Walker, 2007). Later, both extroverted and introverted characters based on the model were implemented in a dialogue system (Andrews, 2012).

When characters are defined based on categories such as age, gender and personality, the variety of characters are then limited: In the real world, two strongly extroverted females in their 20s may have unique qualities that distinguish them, but the system can generate the same character to represent both of them when the system is created using categorical rules.

Dialogue characterization can be achieved by modifying the utterances created by generic dialogue systems using a characterization module. However, a trade-off occurs between the significance of characters in dialogue systems and the quality of modified utterances. When utterances are overly modified, they can be ungrammatical or awkward. To avoid such over-modification, we can use expressions that do not contain concrete semantic implications. An example of this type of expression is back-channelling (BC). BCs are "the short utterances produced by one participant in a conversation while the other is talking" (Ward and Tsukahara, 2000). Although BC has been extensively researched as a means of characterization, especially in spoken dialogue systems , the manner in which BC can affect interactions in text-based dialogue systems is relatively unknown. In our study, we examine the effective use of BCs in text-based dialogue systems for characterising the systems.

2. Corpus Analyses

2.1. Corpus

We used a transcribed corpus of Japanese natural conversation[1] for the analysis of differences in the use of BCs. The corpus consists of Japanese spoken dialogue transcripts. Most dialogues in the corpus are dyadic; however, the corpus also contains some multi-party dialogues. Since the utterers of BCs in multi-party dialogue transcripts cannot be identified, we used only the dyadic dialogues. A summary of the used corpus is given in Table 1. In the transcripts, the BCs were marked with parentheses as shown in Table 2. For example, in the last section of the table, while the participant F140 spoke, the listener (F024 in this case) generated a BC "Jee" which is marked within parenthesis. The other "Jee" utterance is not marked within parenthesis because it was generated by the speaker F140 herself. IDs were assigned to speakers in the corpus. For the male speakers, IDs began with the character "M" followed by a three-digit number. For the female speakers, IDs began with the character "F" followed by a three-digit number. We use this notation to indicate speakers as shown in Table 2. The corpus contained the speakers' attributes, including gender, age, and home town. We could use the set

[1] https://nknet.ninjal.ac.jp/nuc/templates/nuc.html (in Japanese)

of dialogues from one speaker. Also, we could create a set of dialogue for speakers with similar backgrounds in terms of their age, gender and home town.

Attribute	Value
Recording period	Oct. 2001 to Feb. 2003
Number of participants	138 (teenagers to elderlies)
Gender distribution	20 males, 118 females
Dialogue duration	30 to 60 minutes
Dialogue content	Small talk
Number of dialogues	96
Total duration of dialogues	71 hours and 50 minutes
Total number of turns	93,869
Total number of BCs	28,517

Table 1: Summary of the corpus.

F024:	Mr. A has left this.
	[Laughter] Is this a AAA-sized battery?
	(Yes). Let's make some coffee.
F140:	We've been drinking countless cups of black coffee for quite a while.
	[Laughter]
F024:	[Laughter] This is bad for our stomachs.
F140:	But I didn't make it very strong.
	Jee. Perhaps we don't have any more hot water.
	(Jee). Is that so? We have to boil some water first.

Table 2: Excerpts from the dialogue corpus (Translation from Japanese text). The utterances in parenthesis, ("Yes") and ("Jee"), are examples of BCs in the corpus.

2.2. BC Frequency Analysis

We compared the frequencies of BCs of the speakers. The occurrences of BCs per minute were counted for each speaker over all involving dialogues. The results are shown in Figure 1. The three speakers who are used in the later experiments were marked with dotted lines. The average BC frequency is 2.9 per minute. The most frequent occurrence of BCs was 14 times per minute by speaker F118, and the least frequent occurrence was 0.2 times per minutes by speaker F015. The frequency values differed at most 70 times between speakers and we consider that this feature is usable to differentiate characters.

2.3. BC Type Analysis

For classification of BCs, we use the types of BCs as defined in (Den et al., 2012). The types are explained in Table 3. BCs used by the speakers were counted for each BC type. Altogether, the most typical BC type was basic interjection (B), followed by motive interjection (E) and Linguistic response (L). Other types of BCs account for approximately 1 to 2 percent. Then, according to the result of the individual analysis, the ratios of BC types wares quite different from one speaker to another. One speaker consisted of type E BCs more than 60 percent whereas 10% of the utterances of another speaker consisted of the rare type C BCs. Different types of BCs contain different expressions. Therefore, listeners could notice the difference in characters in the different usage patterns of backchannelling types.

BCs not only provide feedback such as agreement/disagreement or understanding to speaker statements, but they offer functions such as acknowledgement and encouragement especially in Japanese which is the language used in our experiment. If BCs are regarded as barge-in responses that elicit turn-taking, and their timing as inappropriate, the utterances may adversely affect users' impression (Hirasawa et al., 1999). Throughout our experiment, the BCs were not considered as barge-ins by the users. BCs in dialogue systems may be evaluated in various aspects, such as human-likeness (Poppe et al., 2013), but we focused here on the characterization.

2.4. Vocabulary Analysis

We considered the vocabularies used by speakers when generating BCs. We searched for words used by less than a certain threshold number of speakers that are considered peculiar to the speakers. We consider that such expression may make the characterization too easy because the user can identify the speaker simply spotting the expression. We experimentally set the threshold value to be five. We excluded nouns, verbs, and adjectives, whose occurrences are dependent on context, from our consideration. Furthermore, words used mostly once were eliminated because they were considered too specific to the particular dialogue in the corpus and hence may not have been reusable in generic dialogues in dialogue systems. As a result, we were able to isolate words used only by the youth, as well as words used by women participants. These words can be used to add characters to the BCs from dialogue systems.

3. Character Generation Method

3.1. Frequency and BC Types

If the occurrences of BCs are not appropriate in terms of timing, people receive negative impression. Therefore, we created a model in order to determine relevant BC placement. The relevance of placement is dependent on the context. Based on the analysis of the corpus, the probabilities of generating BCs were estimated. Since the occurrences of BCs are dependent on the part-of-speech (POS) occurring just prior to the BCs and listener's character, the models were defined as probabilities yielded by the immediately preceding parts-of-speech and the listener. That is, BCs (o) of type j by person (i) are represented by $p_{ij} = P(o_j|POS, i)$. When there are four particles in a dialogue and a type of BC appears once following these, the probability of producing BCs by listeners $P(o|POS_{\mathrm{particle}}, L) = 1/4$. This method places BCs in relevant places in utterances, it cannot infer the ideal timing of BCs. The effect of BC timing is beyond the scope of our present research.

The interaction between the system and users is often short, and BC patterns that are observable in human-to-human dialogues do not evidently appear during the system-to-human dialogue sessions. Since users will not have had

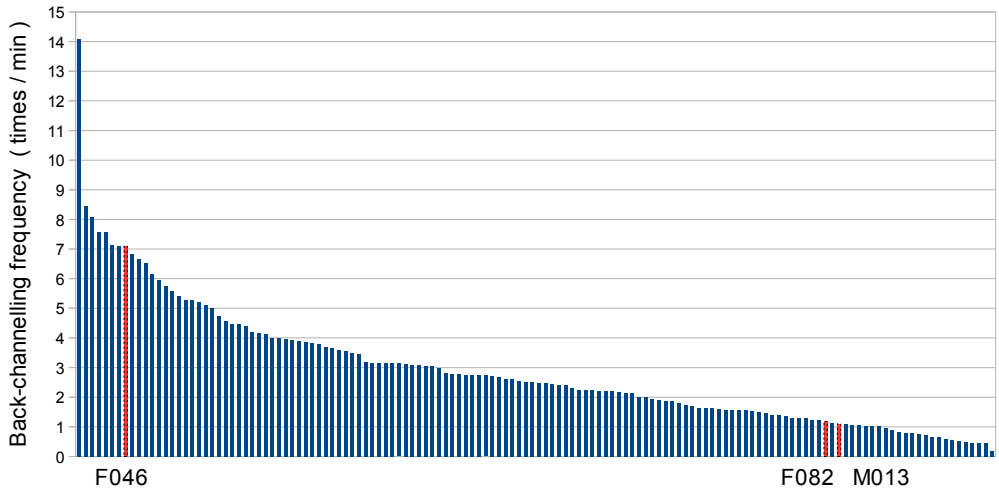

Figure 1: Frequencies of BCs per minute for each speaker.

Type	Description of class	Example BCs
Basic Interjection (B)	Exclamation of acknowledgement or acceptance of messages.	Yes. / Yeah. / Hmmm.
Emotive Interjection (E)	Exclamation that represents a change in mental state such as surprise or awareness.	Oh really? / Aha. Heh.
Linguistic response (L)	Agreement by conventional expressions.	I see. / Indeed. / OK.
Repetition (R)	Repetition of preceding utterances by others.	A: I went to there with a tour group. B:With a tour group.
Complement (C)	Completion of others incomplete utterances by predicting utterance elements	A: Put straw on the shoes, straw rope (B: tie) and tie.
Assessment response (A)	Response to utterances by others with evaluative vocabulary.	Funny. / It's scary.
Other (O)	Responses (back-channels and non-verbal actions) not included in the aforementioned categories.	Really?, . . . , [laugh]

Table 3: Classification of backchannels and examples.

previous exposure to the person represented by the system, they may find it difficult to glean information regarding the characters if the calculated probability is used as is. Therefore, we exaggerated the difference between source speakers. For the exaggeration, we use the multiplication of the probabilities by the deviation. For the set of numbers X and its average $\bar{X}$, the deviation d is defined as $d = x_i - \bar{X}$. With the multiplication parameter α, the probability is changed as follows:

$$p'_{ij} = p_{ij} + (\alpha - 1)(p_{ij} - \frac{\sum_{k=1}^{K} p_{kj}}{|K|}) \qquad (1)$$

where K is the number of speakers considered. This multiplication increases the variance, but does not change the average and the sum of the probabilities, and maintains the order of probabilities among speakers. If the conversion rendered the probability of an occurrence less than 0, the value was reset to 0. We experimentally determined $\alpha = 3$ for the corpus.

3.2. Word Selection

By selecting words less frequently used by listeners than others, characters can be represented. As a measure of the rareness of words, we used word inverse-document frequency (idf). Among several definitions of idf values, we used the following equation:

$$\text{idf} = \frac{|D|}{|D_w|} \qquad (1 \leq \text{idf} \leq |D|) \qquad (2)$$

where $|D|$ is the number of documents or utterances, and $|D_w|$ represents the number of documents or utterances containing the word w.

4. Reactive Dialogue System

We implemented a reactive dialogue system that generated only back-channels with characters. This limited dialogue system is not intended for real application, but was used to test its ability to characterize dialogue by means of only BCs. In conversations, BCs are often generated while a speaker speaks. In text-based dialogue systems, we consider that the users are taking turn when they are inputting text and not completed the utterances. The generation of BCs are incrementally determined for each input morpheme, but sub-utterance units are not considered (Hastie et al., 2013). In our experimental system, while users were taking their turns, the system generated characterized BCs based on probabilities. Values for these probabilities were assigned based on corpus statistics. The probabilities determine types and content of BCs. Expressions used by the system were extracted from the corpus and stored. A BC was displayed, then disappeared before a new BC was shown.

In summary, the system generated BCs in the following manner. The text in the user input space was first morphologically analysed, and the part-of-speech information was obtained. The probability of the back-channelling type was determined based on the $P(o|POS, i)$ as described in Section 3.. The probabilities were adjusted based on the method expressed by Equation 1. Following this, the occurrences of BCs were determined based on the probabilities. If BCs were going to be generated, the corresponding expressions were retrieved from the corpus. The expression used was determined by the weighting according to Equation 2. Finally, the selected BC was displayed.

5. Experiment

5.1. Characters

In our experiments, we chose three personal characters from 138 speakers in the original corpus. Their personal traits were relatively clearer when the BCs were considered. The first character was that of a female speaker F046 in her late teens and used vocabulary associated with young people. The second character also belonged to a woman, F082, in her early 60s. She was polite, and used expressions/words associated with women. The third was a male speaker, M013, in his late 20s, and used expressions/words associated with young men.

5.2. Gender and Age Identification Test

The first test used to assess the effectiveness of our proposed method for character generation for the dialogue system asked participants if they can determine the gender and age of the dialogue systems with which they interacted.

5.3. Person Identification Test

For the second test, we asked participants to identify a person represented in our dialogue system. As a clue, we showed participants three example transcripts of human-to-human dialogue from the corpus. The transcripts were of speakers of the same gender and age categories. One of transcripts is used as the source of BC characterization. Each participant was asked to select one transcript that appeared similar to the dialogue in the system with which they had just interacted.

5.4. Experimental Procedure

Thirteen volunteers participated in our experiments. All were males and in their 20s and 30s. The participants selected dialogue topics among 30 candidates. A participant selected a topic if the participant could talk about it for approximately five minutes. Participants communicated with the dialogue system by inputting text using a keyboard, and the system responded with BCs. Upon completing the dialogue, participants were asked to answer the questions on administered tests as described in previous sections.

6. Experimental Results

The results of our gender-identification test are shown in Table 4. Participants tended to consider the speakers as males. Results of the age identification test are shown in Table 5. In addition, the distribution of the responses is presented in Table 6. The number of participants who correctly answered the questions is shown in bold, (we considered the original and adjacent age groups as correct answers). Participants tended to infer that the speakers were in younger age groups, and the middle-aged speaker F082 did not look her age in terms of BC patterns. The results of the person-identification test are shown in Table 7. For the young female character, the source identities were estimated with 69% accuracy.

By considering the result whereby the minimum accuracy for the gender identification test was high (0.69), and given that participants could guess the gender of the dialogue systems better than by mere chance (male or female), we think that gender characters can be adequately represented by our back-channelling dialogue systems. With regard to age, the characters of younger generations (F046 and M013) could be guessed with high accuracy values, whereas older character (F082) was not clearly represented in our system. This may have been because word usage among the younger generation is often obvious, whereas the vocabularies used by older people are less deviated from formal, which is also considered as standard, language style.

The results for person-identification test showed that BCs alone can distinguish the specific characters of same gender and age background for limited conditions. We asked the participants to rate the amiableness of characters on an ascending five-point scale. The most frequent answer for F046 was the second point, not amiable; whereas the most frequent answer for F082 and M013 were the fourth point, amiable. Therefore, the distinguishable character in terms of BC patterns may probably have an uncomfortable speaking style.

7. Conclusion

In this study, we reimplemented characters for use in reactive text-based dialogue systems by means of BC patterns. Patterns were statistically extracted from a corpus. We showed that we could create diverse characters from them.

Source person	Accuracy
F046	0.69
F082	0.77
M013	1

Table 4: Result of gender identification test.

Source person	Accuracy
F046	0.92
F082	0
M013	0.77

Table 5: Result of age identification test.

	F046	F082	M013
10s	**3**	0	1
20s	**9**	2	**9**
30s	1	8	**1**
40s	0	3	2
50s	0	**0**	0
60s	0	**0**	0
70s	0	0	0
80s	0	0	0
90s	0	0	0

Table 6: Result of age identification test.

Source person	Accuracy
F046	0.69
F082	0.46
M013	0.38

Table 7: Result of person identification test.

Our goal was to provide system users with the feeling that they are interacting with distinct individuals rather than artificially created characters. Through our experiments, we found that some characters can be represented in dialogue systems by statistically extracted BC patterns even among the same age and gender groups.

In future research, we intend to integrate our back-channelling system with a generic dialogue system that can verbally respond to users. Following this, we should be able to compare characters represented through back-channels and regular verbal responses.

8. Bibliographical References

Andrews, P. Y. (2012). System personality and persuasion in human-computer dialogue. *ACM Trans. Interact. Intell. Syst.*, 2(2):1–27.

Den, Y., Koiso, H., Takanashi, K., and Yoshida, N. (2012). Annotation of response tokens and their triggering expressions in Japanese multi-party conversations. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey, May.

Fujie, S., Fukushima, K., and Kobayashi, T. (2004). A conversation robot with back-channel feedback function based on linguistic and nonlinguistic information. In *Proc. ICARA Int. Conference on Autonomous Robots and Agents*, pages 379–384.

Hastie, H., Aufaure, M.-A., Alexopoulos, P., Cuayáhuitl, H., Dethlefs, N., Gasic, M., Henderson, J., Lemon, O., Liu, X., Mika, P., et al. (2013). Demonstration of the PARLANCE system: a data-driven, incremental, spoken dialogue system for interactive search. In *SIGDIAL*.

Hirasawa, J., Nakano, M., and Takeshi Kawabata, K. A. (1999). Effects of system barge-in responses on user impressions. In *Proc. Eurospeech*, pages 1391–1394.

Mairesse, F. and Walker, M. (2007). Personage: Personality generation for dialogue. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 496–503.

Nass, C., Moon, Y., Fogg, B. J., Reeves, B., and Dryer, C. (1995). Can computer personalities be human personalities? In *Conference Companion on Human Factors in Computing Systems*, volume 43 of *CHI '95*, pages 228–229. ACM.

Poppe, R., Truong, K. P., and Heylen, D. (2013). Perceptual evaluation of backchannel strategies for artificial listeners. *Autonomous agents and multi-agent systems*, 27(2):235–253.

Ward, N. and Tsukahara, W. (2000). Prosodic features which cue back-channel responses in english and japanese. *Journal of pragmatics*, 32(8):1177–1207.