Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020), pages 696–701
Marseille, 11–16 May 2020

Relation between Degree of Empathy for Narrative Speech and Type of Responsive Utterance in Attentive Listening

Koichiro Ito[†1], Masaki Murata[†2], Tomohiro Ohno[†3], Shigeki Matsubara[†4]
[†1]Graduate School of Informatics, Nagoya University
[†2]National Institute of Technology, Toyota College
[†3]Graduate School of Advanced Science and Technology, Tokyo Denki University
[†4]Information & Communications, Nagoya University
[†1,4]Furo-cho, Chikusa-ku, Nagoya, 464-8601, Japan
[†2]Eisei-cho 2-1, Toyota, 471-8525, Japan
[†3]Senjuasahi-cho 5, Adachi-ku, Tokyo,120-8551, Japan
ito.koichiro@a.mbox.nagoya-u.ac.jp, murata@toyota-ct.ac.jp, ohno@mail.dendai.ac.jp, matubara@nagoya-u.jp

Abstract

Nowadays, spoken dialogue agents such as communication robots and smart speakers listen to narratives of humans. In order for such an agent to be recognized as a listener of narratives and convey the attitude of attentive listening, it is necessary to generate responsive utterances. Moreover, responsive utterances can express empathy to narratives and showing an appropriate degree of empathy to narratives is significant for enhancing speaker's motivation. The degree of empathy shown by responsive utterances is thought to depend on their type. However, the relation between responsive utterances and degrees of the empathy has not been explored yet. This paper describes the classification of responsive utterances based on the degree of empathy in order to explain that relation. In this research, responsive utterances are classified into five levels based on the effect of utterances and literature on attentive listening. Quantitative evaluations using 37,995 responsive utterances showed the appropriateness of the proposed classification.

Keywords: communication robot, narrative, spoken dialogue, spoken language resource

1. Introduction

To narrate is a fundamental need of human beings. The act of narrating is established only when there is a listener. In the modern society, there are more and more situations when there is no listener and an important issue is how to increase opportunities for people to narrate.

To solve this issue, spoken dialogue agents such as communication robots and smart speakers can play the role of listeners of narratives. In order for such an agent to be recognized as a listener of narratives, it is necessary to convey that it attentively listens to the speaker's narrative. An effective explicit means of realizing this function is to respond to narratives, by generating gestures or utterances. Hereinafter, an utterance used for this function, that is an utterance that responds to the narratives for the purpose of showing an attitude of attentive listening, is simply called *attentive listening response*.

Attentive listening responses show empathy to narrative speech and enhance speaker's motivation to speak. A representative attentive listening response is a back-channel, whose generation methods have already been proposed (Kamiya et al., 2010; Yamaguchi et al., 2016). Besides the back-channel, there are various types of attentive listening responses such as admiration response or evaluation response, and there are some works on generating such responses (Kobayashi et al., 2010; Lala et al., 2017; Meguro et al., 2011; Shitaoka et al., 2017). The degree of empathy shown by attentive listening responses are thought to depend on their type. Empathy to narratives encourages a speaker to speak more only when the degree of the empathy is appropriate. On the other hand, when the degree of the empathy is not appropriate for the narrative, such empathy discourages the speaker. In order to enhance the speaker's motivation to speak, it is necessary to utter an attentive listening response which can show the appropriate degree of empathy. Nevertheless, the relation between types of attentive listening responses and degrees of empathy has not been explored.

This paper describes the classification of types of attentive listening responses based on the degree of empathy to narratives, in order to explain this relation. This research analyzes the degree of empathy to narratives expressed by attentive listening responses, considering the effects of responses and literature on attentive listening, and then classifies the types of attentive listening responses into five levels. The appropriateness of this classification is quantitatively evaluated using attentive listening response data comprised of responsive utterances to narrative speech (Ohno et al., 2017).

This paper is organized into five sections. Section 2 explains attentive listening responses and degree of empathy shown by them. In Section 3, types of attentive listening responses are classified into five levels, while Section 4 provides the results of quantitative evaluation of the appropriateness of the proposed classification. Finally, in Section 5 summarizes the paper.

2. Degree of empathy and types of attentive listening responses

Attentive listening responses show empathy to narratives and enhance speaker's motivation to speak. Back-channel feedback is a representative attentive listening response, but besides it there are various types of attentive listening responses such as admiration or evaluation. Figure 1 shows an example of narrative speech and attentive listening responses. Here strings in parentheses show the types of the

Narrative Speech		Attentive listening response	
イタリア旅行をしたことが	I enjoyed traveling	はい	yes (back-channel)
一番楽しかったです	to Italy the most	はーそうですかー	is it so? (admiration)
		素敵ですねー	it is wonderful (evaluation)
		ふーん	hmm (admiration)
もう二度と行かないかなと	I went there thinking	イタリア旅行	Traveling to Italy (echoic response)
思いながら行ってきましたけど	that I can not go again	いえいえそんなー	That's not true (disapproval)
		あーそうですかー	oh, is it? (back-channel)

Figure 1: Example of narrative speech and attentive listening response

Types of response	Roles, ie. what does the response show?
back-channel	hearing success
admiration	attitude of admiration, surprise, or attention to the content of the speaker's utterance
evaluation	attitude toward the situation described by speaker's utterance
approval	attitude of approval of the content of the speaker's utterance
disapproval	attitude of disapproval of the content of the speaker's utterance
echoic response	comprehension of the content of the speaker's utterance and a sense of security
paraphrasing	attitude of trying to understand and share the content of the speaker's utterance
satisfaction	listener's attitude that the content of the speaker's utterance is satisfactory for him/her
surprise	attitude of strong surprise at the content of the speaker's utterance
surprise with doubt	attitude of surprise or doubt toward the content of the speaker's utterance
opinion	listener's personal experiences, opinions, or feelings
complement	listener's status that he/she is eagerly listening to the speaker's utterance
greeting	acknowledgement of the speaker's presence and willingness to favorably interact with the speaker
provoke memory	listener's reaction that his/her memory is provoked by the content of the speaker's utterance
start thinking	listener's reaction that he/she is starting to think about the content of the speaker's utterance
thinking process	listener's status that he/she is thinking about the content of the speaker's utterance

Table 1: Types of attentive listening responses and their roles

attentive listening responses. The degree of empathy expressed by attentive listening responses is thought to depend on their types. Here is an example of narrative and its two attentive listening responses:

[narrative] I also like calligraphy and have received the Prime Minister Prize

[response 1 (back-channel)] yeah

[response 2 (evaluation)] that's amazing

Here, strings in parentheses show the types of responses. The degree of empathy to the above narrative expressed by "response 2" is thought to be higher than the one by "response 1." It is important to explore the relation between the types of attentive listening responses and degree of empathy because providing responses with an appropriate degree of empathy to narratives effectively encourages a speaker to speak more. This study classifies the types of attentive listening responses based on the degree of empathy shown to narratives in order to explain this relation.

3. Empathy level classification of types of attentive listening responses

This research examines 16 types of attentive listening responses in attentive listening response data (Ohno et al., 2017). These types are defined based on literature (*Nihongo Kijutsu Bunpo Kenkyukai*, 2009) about attentive listening responses. According to literature, what attentive listening responses show to narratives depends on their types. Table 1 shows the target types of attentive listening responses and what they show to narratives.

The 16 types of attentive listening responses are classified into five levels based on the degree of empathy estimated from their roles in Table 1, which is shown in Figure 2. The number after "Level" indicates the degree of empathy, whereby the degree of empathy is the lowest in responses on level 1, where the back-channel belongs, and it is the highest in responses on level 5, where opinion belongs. Here is an example of back-channel and opinion to a narrative:

[narrative] it is the happiest for me to be blessed with health

[response 3 (back-channel)] yes

[response 4 (opinion)] being healthy is the best

The degree of empathy shown to the above narrative by "response 4" is thought to be much higher than the one by "response 3."

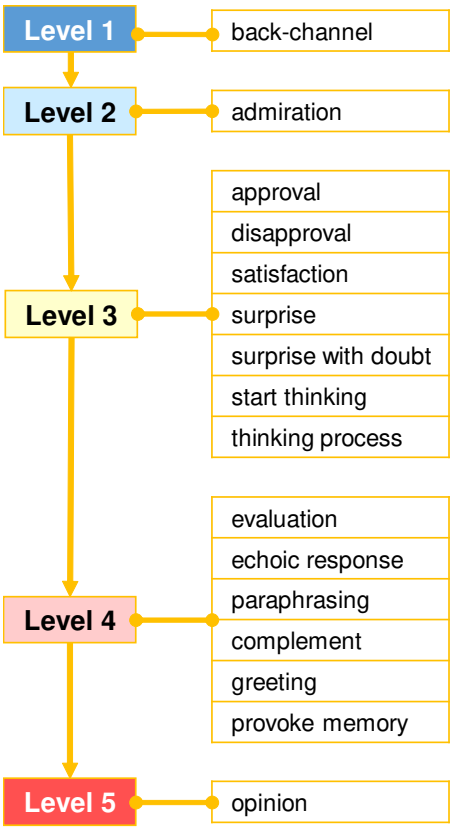

Figure 2: Classification of types of attentive listening responses

4. Evaluation of appropriateness of empathy level classification

The 16 types of attentive listening responses were classified into five levels in Section 3. The classification was made based on the degree of empathy estimated by considering the response roles in Table 1. This means that the classification was made qualitatively. Therefore, it is not certain whether or not the proposed classification is appropriate. In this section, the appropriateness of proposed classification is quantitatively evaluated using attentive listening response data (Ohno et al., 2017). This data contains attentive listening responses given by five listeners to the same narrative speech. The responses were collected in the following way: five listeners respectively respond to narrative speech in Japanese Elder's Language Index Corpus (JELiCo)[1] at an arbitrary timing while listening to the speech. Figure 3 shows an example of the narrative speech and an attentive listening response by one of the five listeners. In the three columns about narrative speech, each row means information on a morpheme or a pause. In the four columns about the attentive listening response, each row means information on the response to the narrative speech. As mentioned above, this data has five listeners' responses to the same narrative speech. That is, there are four more attentive listening responses in Figure 3.

This section quantitatively evaluates the appropriateness of the proposed classification, using a part of attentive listening response data. Hereinafter, a part of this data used for evaluation is simply called evaluation data. Figure 4 shows the breakdown of responses in evaluation data. Evaluation data has 37,995 attentive listening responses in total. It is confirmed that responses on level 1 account for more than 65% of all responses and those in either level 1 or 2 account for more than 85%.

4.1. Evaluation in terms of concreteness

It was evaluated whether the classification of types of attentive listening responses is appropriate in terms of response concreteness. In this section, the concreteness of the response refers to its information value. It is considered that the concreteness of the response depends on their type. Here is an example of a narrative and its two attentive listening responses:

[narrative] I enjoyed traveling to Italy the most
[response 5 (admiration)] uh
[response 6 (echoic response)] Italy travel

The concreteness of "response 6" is higher than that of "response 5." Furthermore, the degree of empathy shown by "response 6" is also thought to be higher than that by "response 5." Since for uttering responses with the high concreteness like "response 6" it is necessary to deeply consider the content of narratives, the degree of empathy shown by these responses tends to be high. On the other hand, since it is not necessary to deeply consider the content of narratives for uttering responses with low concreteness like "response 5," the degree of empathy shown by these responses tends to be low. Therefore, the concreteness of responses is thought to reflect the degree of empathy they express.

This research measured the concreteness index of responses. The following two features were adopted as the index:

1. *length*: the number of morphemes in the response
2. *info*: information value of the response

The second feature *info* is defined as the summation of information value of all content words contained in the response. In this research, $info(r)$, information value of the response r, is measured using the following equation:

$$info(r) = \sum_{w \in (r \cap W_C)} \log_2 \frac{\Sigma_{w' \in (CSJ \cap W_C)} F(w')}{F(w) + 1} \quad (1)$$

where W_C is a set of content words, w is a content word contained in a response r, $F(w)$ is the frequency of w in the Corpus of Spontaneous Japanese (CSJ) frequency list of short unit words ver. 2018.3.1[2], and $\Sigma_{w' \in (CSJ \cap W_C)} F(w')$ is 3,325,907, which is the sum of the number of content words in CSJ frequency list. We measured these two feature values of all responses in evaluation data and then calculated the averages per response type. Finally, we calculated the averages per level using those per response type.

[1] https://www.gsk.or.jp/catalog/gsk2018-a

[2] https://pj.ninjal.ac.jp/corpus_center/csj/chunagon.html

Narrative speech			Attentive listening response			
start - end	surface form	English translation	start - end	response	English translation	response type
0.03 - 0.45	自分	my				
0.45 - 0.53	の	-				
0.53 - 0.88	趣味	hobby				
0.88 - 1.26	は	is				
1.26 - 2.27	pause	-	1.18 - 1.46	はい	yes	back-channel
2.27 - 2.67	社交	ballroom				
2.67 - 3.02	ダンス	dancing				
3.02 - 3.31	です	-	3.40 - 3.58	あっ	ack	admiration
3.31 - 5.82	pause	-	3.91 - 4.06	いいですね	sounds good	evaluation

[narrative] My hobby is ballroom dancing

Figure 3: Example of attentive listening response data

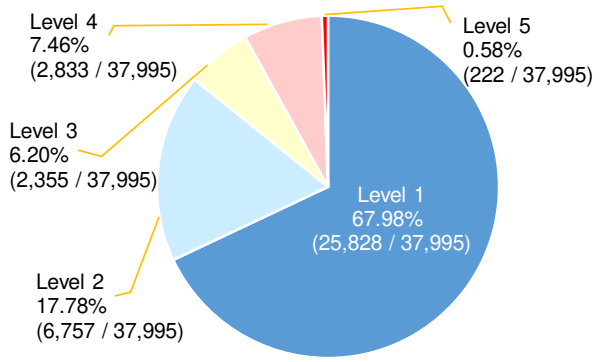

Figure 4: Breakdown of responses in evaluation data

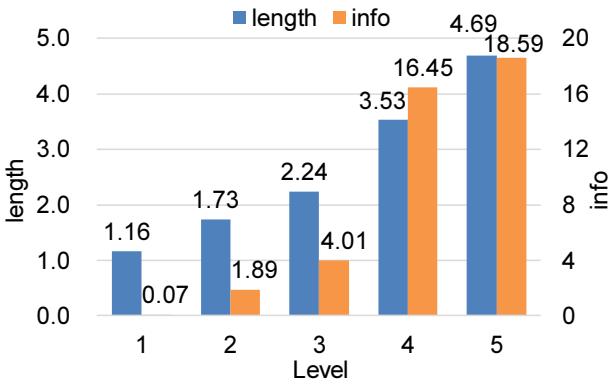

Figure 5: Concreteness of attentive listening responses

Figure 5 shows these results. It is confirmed that the higher the level is, the higher both *length* and *info* are. This indicates that the classification result in Section 3 is appropriate. Furthermore, the difference between *info* of level 3 and that of level 4 is much larger than the difference between the other two pairs. This indicates that the difference between the degree of empathy shown by responses on level 3 and that by those on level 4 is also very large.

4.2. Evaluation in terms of versatility

It was evaluated whether the classification of attentive listening responses is appropriate in terms of response versatility. For example, echoic responses can be uttered only when there are phrases that deserve to be repeated in narrative, while evaluation responses can be uttered only when there are contents that deserve to be evaluated in narrative. There are not many phrases and contents like these in narrative. Therefore, the timings when these responses can be uttered are limited and the versatility of these responses is low. For uttering these responses in appropriate timings, it is necessary to deeply understand the content of narrative. Therefore, contrary to the versatility, the degree of empathy shown by these responses to narrative tends to become high.

Admiration responses can be uttered only when there are contents that deserve to be admired in narrative, and there are more contents like these in narrative than phrases or contents that deserve to be repeated or evaluated. Furthermore, it is considered that back-channel responses can be uttered regardless of contents of narrative. Therefore, the timings when these responses can be uttered are diverse and the versatility of these responses is high. For uttering these responses in appropriate timings, it is not necessary to deeply understand the content of narrative. Therefore, contrary to the versatility, degree of empathy shown by these responses to narrative tends to become low. That is, the versatility and degree of empathy of the response are in a trade-off relation.

In this section, we used evaluation data which has responses given by five listeners for this evaluation. The versatility of responses is defined based on their co-occurrence. Since responses of high versatility are uttered at various points in narrative speech, it is considered that these responses occur along with many other types of responses. On the other hand, since responses of low versatility are uttered in fewer points in narrative speech, it is considered that these responses do not occur along with many other types of responses. This means that the versatility of the response is the variety of responses occurring along with a response.

narrative speech	listener's responses					co-occurrence with はい
	listener 1	listener 2	listener 3	listener 4	listener 5	
趣味 と いっ ても	はい	はい	うん	はい		うん
今 続け て いる		うん	ええ			
こと では ない けれど	ええ	はい	はい	ええ	はい	ええ
着付け を し たり	はい	あ / はいはい	はい	着付け	あっ	あ / はいはい / 着付け/ あっ
編み物 を し たり	へー	はい	うーん / はい	ええ	へえ	へー / うーん / ええ / へぇ

[narrative] I do not continue now, but my hobbies are dressing, knitting and

(1) $F_{hai}(u) = 2\,(u = ee),\ 1\,(u \in \{un, a, haihai, kitsuke, a', \overline{he},\ \bar{u}n, hee\}),\ 0\,(otherwise)$

(2) $ent(hai) = -\frac{2}{10}\log_2\frac{2}{10} - 8(\frac{1}{10}\log_2\frac{1}{10}) = 3.12$

Figure 6: Example of measuring $ent(r)$, versatility of the attentive listening response r. This example shows how to measure $ent(hai)$, versatility of the response *hai*: (1) calculate $F_{hai}(u)$, a co-occurrence frequency between *hai* and another response u ($\neq$ *hai*); (2) measure $ent(hai)$ using $F_{hai}(u)$.

In this research, $ent(r)$, the versatility of surface form of response r, is defined using the following equation:

$$ent(r) = -\sum_{u} \frac{F_r(u)}{\sum_{u'} F_r(u')} \log_2 \frac{F_r(u)}{\sum_{u'} F_r(u')} \qquad (2)$$

where $F_r(u)$ is a co-occurrence frequency between r and another response u ($\neq r$), that is, the number of timings when both r and u exist in the responses uttered by the five listeners.

This section evaluates whether the classification presented in Section 3 is appropriate, using value of ent. For measuring ent, information on co-occurrence of responses is necessary. Therefore, the start time of response was mapped onto the nearest position of the following one in the narrative speech, considering that responses tend to be uttered right after a linguistic or phonetic boundary, namely:

1. clause boundary (linguistic)
2. pause longer than 200 milliseconds (phonetic)

Evaluation data has 10,469 boundaries in total. Hereinafter, this mapped position of response start time is simply called response timing. In this research, we defined co-occurrence responses as those with the same response timing.

Figure 6 shows an example of measuring ent and also shows how to measure $ent(hai)$, that is, the ent of response *hai*. Narrative speech and responses to the speech are shown in the upper part of this figure, and a way to measure $ent(hai)$ is shown in the lower part. In the "narrative speech" column, each row refers to a segment of narrative speech by considering response timing. In the "listener's responses" column, each row means responses whose response timing is right after the narrative speech segment in the same row. The "listener's responses" column is divided into five more columns because there are responses by five listeners in the evaluation data. In the "narrative speech" and "listener's responses" columns, each row has three more rows. The first two rows refer to the Japanese surface form and pronunciation of narrative speech or response, and the third row contains their English translation. Each row in the last "co-occurrence with *hai*" column contains the responses of responses in the row that co-occurred with *hai*.

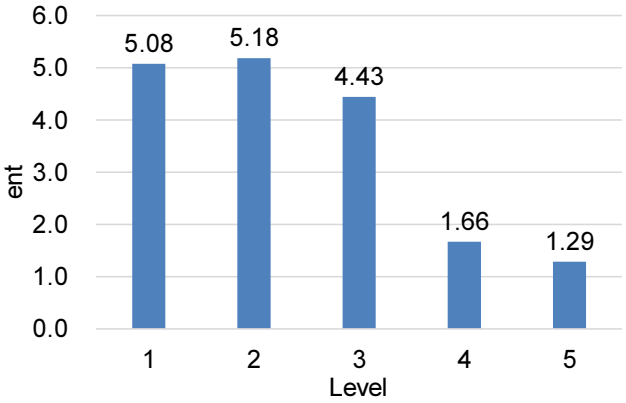

Figure 7: Versatility of attentive listening responses

We measured ent of all responses in evaluation data and then calculated the averages per response type. Finally, we calculated the averages per level using those per response type. Figure 7 shows this result. Except from the fact that ent on level 1 is little lower than that on level 2, it is confirmed that when the level is lower, ent is higher. This result confirms that the classification proposed in Section 3 is appropriate. Furthermore, the difference between ent on level 3 and that on level 4 is much larger than the difference between the other two pairs. This indicates that the difference between the degree of empathy shown by responses on level 3 and that by those on level 4 is very large.

5. Conclusion

This research classified the types of responsive utterances that show that the listener is attentive to narrative speech using the degree of empathy as the criterion. It evaluated the appropriateness of the proposed classification using attentive listening response data. In the future, we will develop a spoken dialogue agent to generate responses showing an appropriate degree of empathy to narratives.

6. Acknowledgments

The narrative corpus was provided by the Social Computing Laboratory of Nara Institute of Science and Technology.

This research was in part supported by the Grand-in-Aid for Challenging Exploratory Research (No. 18K19811).

7. Bibliographical References

Kamiya, Y., Ohno, T., and Matsubara, S. (2010). Coherent back-channel feedback tagging of in-car spoken dialogue corpus. In *Proceedings of the 11th Annual SIGdial Meeting on Discourse and Dialogue (SIGDIAL-2010)*, pages 205–208.

Kobayashi, Y., Yamamoto, D., Koga, T., Yokoyama, S., and Doi, M. (2010). Design targeting voice interface robot capable of active listening. In *Proceedings of the 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI-2010)*, pages 161–162.

Lala, D., Milhorat, P., Inoue, K., Ishida, M., Takanashi, K., and Kawahara, T. (2017). Attentive listening system with backchanneling, response generation and flexible turn-taking. In *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue (SIGDIAL-2017)*, pages 127–136.

Meguro, T., Higashinaka, R., Minami, Y., and Dohsaka, K. (2011). Evaluation of listening-oriented dialogue control rules based on the analysis of HMMs. In *Proceedings of the 12th Annual Conference of the International Speech Communication Association (Interspeech-2011)*, pages 809–812.

Nihongo Kijutsu Bunpo Kenkyukai, editor, (2009). *Gendai nihongo bunpo 7*, pages 165–182. *Kuroshio Shuppan*. (In Japanese).

Shitaoka, K., Tokuhisa, R., Yoshimura, T., and Hoshino, H. (2017). Active listening system for a conversation robot. *Journal of Natural Language Processing*, 24(1):3–47. (In Japanese).

Yamaguchi, T., Inoue, K. Yoshino, K., Takanashi, K., Ward, N. G., and Kawahara, T. (2016). Analysis and prediction of morphological patterns of backchannels for attentive listening agents. In *Proceedings of the 7th International Workshop on Spoken Dialogue Systems (IWSDS-2016)*, pages 1–12.

8. Language Resource References

Ohno, T., Murata, M., and Matsubara, S. (2017). Collection of responsive utterances to show attentive hearing attitude to speakers. In *Proceedings of the 11th International Conference on Ubiquitous Information Management and Communication (ACM IMCOM-2017)*, pages 1–4.