PAPER *Special Section on Natural Language Processing and its Applications*

Activating Humans with Humor —— A Dialogue System That Users Want to Interact with

Pawel DYBALA[†a)], **Michal PTASZYNSKI**[†], ***Nonmembers***, **Rafal RZEPKA**[†], ***and*** **Kenji ARAKI**[†], ***Members***

SUMMARY The topic of Human Computer Interaction (HCI) has been gathering more and more scientific attention of late. A very important, but often undervalued area in this field is human engagement. That is, a person's commitment to take part in and continue the interaction. In this paper we describe work on a humor-equipped casual conversational system (chatterbot) and investigate the effect of humor on a user's engagement in the conversation. A group of users was made to converse with two systems: one with and one without humor. The chat logs were then analyzed using an emotive analysis system to check user reactions and attitudes towards each system. Results were projected on Russell's two-dimensional emotiveness space to evaluate the positivity/negativity and activation/deactivation of these emotions. This analysis indicated emotions elicited by the humor-equipped system were more positively active and less negatively active than by the system without humor. The implications of results and relation between them and user engagement in the conversation are discussed. We also propose a distinction between positive and negative engagement.

key words: *Human-Computer Interaction, dialogue, AI and social sciences, psycholinguistics*

1. Introduction

The field of Human-Computer Interaction (HCI) is recently gathering more and more attention from AI researchers. Some of them focus on conversational aspects of HCI, aiming to construct a system able to converse with its users. Some areas, however, are still to be explored — one among them concerning users' broadly-defined "engagement" in the interaction. This paper contributes to this field. Based on experimental data from our previous works [1], [2], we analyze results of evaluation experiments of two (humorous and non-humorous) non-task oriented conversational systems in order to check users' engagement. The role of humor as a means to enhance the degree of user involvement in the interaction is also investigated. Lastly, we present the novel approach of making a distinction between positive and negative engagement.

The results described here are important and useful in several ways. First of all, the mechanisms of human engagement in HCI are still not thoroughly investigated. If methods of making users invest in these interactions with computers are found, better, more user-friendly systems can be constructed. Such investment is of high importance in applications of natural language processing systems such as vehicle navigation systems (i.e., talking to keep drivers awake) or educational software. In our research, we point to humor as a contributing factor to positive engagement. This stands to yield very practical knowledge. The role of humor in human-friendly systems should not be underestimated.

1.1 Computers That Chat

Despite attempts to euphemize "chatterbox" systems as "casual" or "free talking conversational systems", the former best describes what they do: they chat. Chatterbots are often seen as opposition to task-oriented systems (e.g. information kiosks, tour guiding agents) which aim to achieve well-specified goals. Despite this contrast, to say the only goal of a chatterbot is amusement would be incorrect. Of course we humans do chat for pleasure, but we also do so to maintain social relations, or, simply, to feel that we are not alone. However, in modern times our companions are often made of silicon and wires. In fact, according to SRCT (Social Response to Communication Technologies) theory [3], we tend to treat computers as they were social actors, in a similar way we treat other humans. Therefore, it is desirable that they would be able not only to perform a strictly task-oriented conversation, but also converse with us.

Such ability could fundamentally be implemented in any machine that interacts with humans. There are, however, some fields in which a freely talking system would be highly beneficial. Such systems include vehicle navigation systems that work to keep drivers awake through conversation, and conversation partners for lonely elderly people.

1.2 Engagement in HCI

Even if we succeed to build a talking engine able of generating perfectly correct utterances, it will mean nothing if users are unwilling to interact with it. Therefore, it is advisable to take into consideration also such factors as users engagement in the conversation and interaction in general.

Although the term "engagement" is quite difficult to define, for the needs of this research we define it as the degree to which a user is taking part or willing to continue a conversation/interaction.

Albeit being rather neglected, (especially in relation with chatterbots), engagement in HCI was the subject of numerous research projects, most of which focused on the interactions between humans and robots, and, subsequently, on non-verbal (physical) aspects of engagement, such as

Manuscript received March 20, 2009.
Manuscript revised July 9, 2009.
[†]The authors are with the Graduate School of Information Science and Technology, Hokkaido University, Sapporo-shi, 060–0814 Japan.
a) E-mail: paweldybala@media.eng.hokudai.ac.jp
DOI: 10.1587/transinf.E92.D.2394

gestures and gaze [4]. Although this approach is also of high importance (according to existing research as much as 65% of human-human communication is nonverbal), the role of verbal interaction must not be forgotten.

The correlation between a speaker's emotiveness and conversational engagement has been investigated [5]. Hall et al. [6], for example, studied the role of so called "empathic engagement" with synthetic characters in a virtual learning environments for kids. The scope of this empathic engagement, however, is slightly different from what we are investigating in this study, as it referred mostly to the children's compassion and empathy towards virtual characters.

A more explicit study was conducted by Yu et al. [7]. In this research a user's engagement level was measured based on emotion derived from the user's speech. They proposed a method of detecting user engagement using machine learning, based on the parameters of user emotiveness and arousal level. This method reached an accuracy level of 63%, which may or may not direct proof of causation between emotiveness and engagement, but at least it does show that it is possible to detect one on the basis of another, implying at least some form of relation.

The approach presented by Yu et al. is quite similar to one presented in this paper. In this paper, however, focus is placed on emotions extracted from textual layer of conversation (see Sect. 3.4), the role of engagement in non-task oriented conversation and, primarily, the influence of humor as a measure to improve the engagement.

To our knowledge, there is no existing research concerning the influence of humorous stimuli on user engagement in interaction with chatterbots. Therefore, this paper can be seen as a novel contribution to this field. In addition, Yu's et al. algorithm does not detect specific types of emotions, and there is no distinction between positive and negative engagement.

Unfortunately, to our knowledge, there is no comparative study concerning the role of engagement in conversation with task- and non-task oriented systems. However, in case of the task-oriented systems, engagement and users satisfaction seems to be somewhat easier to achieve. The very existence of a task to be completed together mutually between human and system binds the participants of interaction together and, by definition, elicits user engagement. In a non-task-oriented dialogue, however, without a specific goal in the conversation, user engagement depends more on the content of a partner's utterance.

1.3 Small-Scale Experiment

Although some of existing works on engagement (see Sect. 1.2), especially the one by Yu et al. [7] show that there is a connection between engagement and emotions, we have not found any research that explains the relation between these two features directly. Therefore, we decided to conduct our own small-scale experiment to investigate this subject.

Table 1 Results of small-scale experiment — relation between emotiveness and engagement.

	Which more emotive?	Where more engagement?
Dialogue 1	10.0%	3.3%
Dialogue 2	90.0%	96.7%

From a corpus of Japanese human-human dialogues [8], we chose two conversations, one between two company workers and one between two female high school students. The criterion for choosing these two conversations was that we wanted them to differ in terms of emotiveness. Thus, the dialogue between company workers was assumed to be much less emotive than the one between schoolgirls. To check this assumption, we also included an emotiveness evaluation in the experiment.

Next, we prepared two questionnaires. Each contained the two dialogues and a different set of questions:

Questionnaire 1: **which dialogue was generally more emotive?**

Questionnaire 2: **which dialogue's participants were more engaged in the conversation?**

Each questionnaire was given to 30 evaluators (university students). The results are summarized in Table 1.

As shown in Table 1, the results of this small-scale experiment showed that the conversation assessed as more emotive by 90.0% of evaluators was also assessed as including more engagement (96.7%). As both dialogues were of the same length (18 turns each) and there were no other visible differences between them, it can be stated that the more emotive the dialogue and its participants are, the higher their engagement in the conversation tends to be. Needless to say, the experiment described in this section should be repeated with a bigger set of dialogues; however, we believe that even in the small scale of data the tendency of positive relation between emotions and engagement is visible.

The results are consistent with those presented by Yu et al.

1.4 Engagement and Emotions

Acquiring information about people and their attitudes on the basis of conveyed emotions is not a new idea. One of the best known concepts in this area is the "affect-as-information" approach, first proposed by Schwarz and Clore [9]. The main idea of this approach is based on the claim that humans use affect in the same way as any other criteria; namely by using the informational value of their affective reactions to form opinions and judgments. This leads to the assumption that information about someone's attitude to a product (here — to a conversational system) can be derived from information about changes in his or her affective states during its usage.

Subsequently, the number and type of emotions conveyed by speakers can give us information about their involvement in the conversation. In other words, it can be assumed that the more emotions users show towards the dialogue system, the higher their level of engagement in

the conversation. However, there is also a need to distinguish between types of engagement. Anger, for example, increases the level of engagement, but not in a desirable way. Therefore, we propose distinguishing between positive and negative engagement as associated with positive/negative emotions.

1.5 Humor in Chatterbots

Humor is still not very popular as a subject of research in computational sciences. Very few actually implemented humor-equipped systems exist, and among those, only simple forms of humorous text are processed. Probably the most well-known of these is Binsted's JAPE system [10], capable of generating simple pun-based riddles. Later on, JAPE was experimentally integrated by Loehr into the dialogue system Elmo (designed to act as a player in a text-based virtual game) [11]. The results, however, were not satisfying, as there was no relevance between user utterances and the system's humorous responses. Since that time, to our knowledge, no functional humor-equipped chatterbot have been constructed.

There are few studies on the influence of humor on the quality of interaction in HCI. For example, research conducted by Morkes et al. [12] showed that a humor-equipped (albeit not humor-generating) task-oriented system was evaluated as more sociable, likeable and easier to cooperate with by users. Additionally, in our previous works [1], we showed that the presence of humor in non-task oriented dialogue can signficantly enhance conversational system's performance in the eyes of users.

This feature can be used in such applications as dialogue systems in car navigators. There are robust proofs showing that humorous content activates the same parts of brain activated during pleasurable emotional states induced by eating, listening to enjoyable music or having sex [13]. Therefore, devices used to keep the user awake, the presence of humor to activate the brain would help prevent them from getting bored or falling asleep.

In this research, we focus on puns as they are relatively easy to compute using NLP methods, and as Japanese is rich in homophonic phrases compared to most other languages.

1.6 Our Contribution

The most important novel research in this paper is the investigation of the relation between humor, user engagement, and emotions. Also, to our knowledge, the conversational aspect of engagement has not yet been studied in case of chatterbots. We also propose a distinction between positive and negative engagement.

2. Outline of the Study

2.1 Hypotheses and Expectations

Before conducting the analysis described below, we assumed that the presence of humor would visibly improve positive user engagement in the conversation. Thus, we formed the following sets of hypotheses:

Null hypothesis: User engagement (related to emotiveness as showed in Sect. 1.2 and Sect. 1.3) will be the same in conversations with the system with and without humor.
Alternative hypothesis: User engagement and emotiveness will be higher in conversations in the system with humor.

Null hypothesis: User willingness to continue the conversation will be the same for both systems.
Alternative hypothesis: User willingness to continue the conversation will be significantly higher for the system with humor.

Null hypothesis: The proportion of positive and negative emotions will be the same for both systems.
Alternative hypothesis: Users will convey more positive emotion in conversations with the system with humor than with the system without humor.

We also expected to find differences in the dimension of activation/deactivation. Method to verify the hypotheses is described in Sect. 2.2.

2.2 Method

In order to verify the hypotheses, we:

1) Asked 13 users to perform conversations with two similar systems, one of which was humor-equipped (see Sect. 3.1 and Sect. 3.3);

2) Analyzed the chat logs with emotiveness analysis system (Sect. 3.4) and compared results for both systems.

3) Projected the results on Russel's two-dimensional emotiveness space (see Sect. 3.4);

4) Analyzed the results to compare the degree and type (positive/negative and activated/deactivated) of user engagement for both systems.

Based on the overall results, we discuss the relation between humor, emotions and types of engagement.

3. Systems Used

In this research we conducted comparative evaluation experiments using two dialogue systems: with and without humor. The former was equipped with our pun generating engine (see Sect. 3.3). To analyze the results, we used the ML-Ask emotive analysis system, which recognizes emotions from the utterances (see Sect. 3.4).

3.1 Chatterbot

The first system in our research is a freely talking keyword-based conversational system, created by Higuchi et al. [14]. The conversation topic is chosen freely by the user, and the system extracts related sets of words, based on keywords spotted in the user's utterance. Next, word associations are

extracted in real time using Goo search engine† snippets (using no prepared resources such as off-line databases).

In the next step, the system applies extracted the word associations into proposition templates, like: [(noun) (topic indicating particle *wa*) (adjective)]. Next, the naturalness of a proposition phrase is checked on the Internet. If the proposition is recognized as unnatural (low result frequency in Goo), it is deleted and the system generates a new one in the same way.

Finally, the system adds modality (expressions such as "well" or "yeah,") to the extracted natural proposition and again checks the semantic correctness of the proposed sentence on the Internet.

In summary, the system answers user's utterance with a sentence corresponding the topic with modality. An example of such conversation can be found below:

User: - *Nanika sukina tabemono aru?* (What food do you like?)

System: - *Maa, tabemono-wa oishii desu.* (Well, food tastes good.)

User: - *Saikin-wa osake-mo sukini nattekitanda.* (Recently, I begin to like alcohol too.)

System: - *Demo, sake-wa yowai-no-yo-ne.* (But, I can't drink much.)

The chatterbot was also used as a base for creating humor-equipped system — see Sect. 3.2 and Sect. 3.3 for details.

3.2 Pun-Generator

The pun-generating engine, developed in our previous research [1], is also based on Internet data. The system extracts a base word from a user utterance (usually a noun) and transforms it using Japanese pun phonetic generation patterns to create a phonetic candidate list. Than it checks all candidates in the Goo search engine, and chooses the one with the highest hit rate. Next, it uses the KWIC on WEB — online Keyword-in-context sentences database [15] — to find a sentence with the chosen word and extracts its part starting with the word. Below is an example of the system's process:

User: ***Jaa, issho ni eiga wo mi ni ikanai?***
(So, will you go to see a movie with me?)

Base word: *eiga* (a movie)
Pun candidate: *eiga* (glory)
KWIC Sentence part: ...*eiga wo hokotta* (was glorious)

System's response: ***Eiga* (movie) *to ieba eiga* (glory) *wo hokotta.***
(Speaking of movies, it was really moving!)

If no candidate is found for user's input, the system randomly selects a pun from our pun data base.

3.3 Joking Chatterbot

Two algorithms described above were joined to create a talking system which tells jokes. A very simple rule was applied for joke frequency: every third response was replaced by a sentence including a joke. Following this, every third user utterance the input for the pun generator, which generates an appropriate pun for it. Small-scale tests showed that joking in every third turn is optimal for this experiment. This method, although quite simple, allowed us to check if the usage of humor improved the system's overall performance. [1]

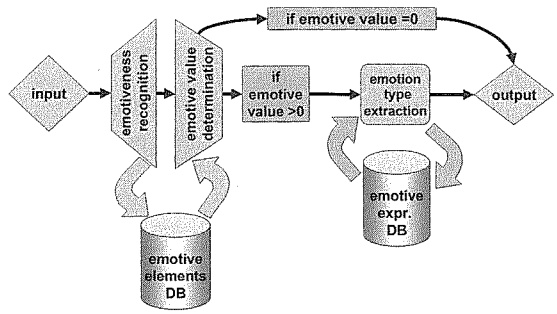


Fig. 1 Flow chart of the ML-Ask system.

3.4 Emotiveness Analysis System

To check what emotions our humor-equipped talking system triggered in users, we used Ptaszynski's et al. ML-Ask Emotive Elements/Emotive Expressions Analysis System for Japanese, which determines an utterance's emotiveness and types of emotions [2]. Based on the ideas presented by Ptaszynski in his earlier publication [16], the system performs utterance analysis in two general steps:

1. Determining general emotiveness (emotive/non-emotive) and

2. Specifying types of emotions (in emotive utterances only).

The system's algorithm outline is presented on Fig. 1.

In the first step, the presence/absence (emotiveness recognition) and amount of emotions (emotive value determination) in user's utterances is checked. For example, the sentence:

***"Kyo wa nante kimochi ii hi nanda!* (Today is such a nice day!)"**,

is recognized as emotive, as it contains emotive expression: *kimochi ii* (nice) and emotive elements: *nante* (such), "*nanda*" (emphasis) and exclamation mark, which do not belong to any particular type of emotions, but make the utterance more emotive.

In step 1, the above sentence is denoted as: emotive, with emotive value = 4 (total number of emotive expressions and elements).

In the second step, the emotive value (number of emotive expressions) of the utterance is checked. If there are emotive values in the utterance, analysis of specific types of emotions is conducted. Knowledge about emotions shown by the evaluators during the conversation provides us with information on their feelings towards the system. If detected emotions were positive or changed from negative through

†Goo search engine, http://www.goo.ne.jp/

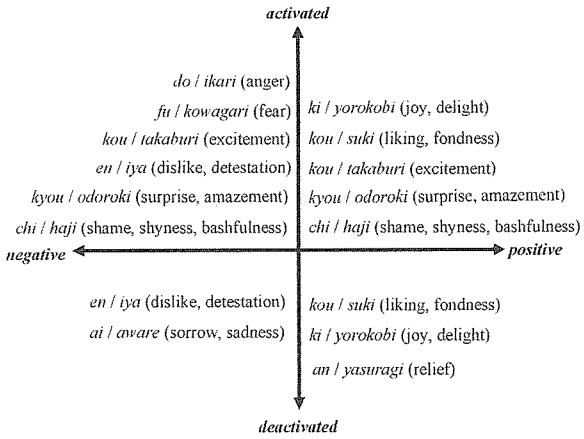


Fig. 2 Grouping Nakamura's classification of emotions on Russell's two-dimensional space.

neutral to positive during the conversation, the general sentiment towards the talking system was considered as positive. If emotions are negative or change from positive through neutral to negative during the conversation, the general sentiment towards the system was considered negative.

In the above example sentence the system finds the emotive expression "*kimochi ii* (nice)", which belongs to the group called "*yorokobi* (joy)". Therefore, the sentence is recognized as: 1) **emotive** and 2) **positive**. In Japanese, emotional engagement in the conversation suggests a tendency to familiarize with the partner [17] — which, in this case, is the talking system.

Division of emotions to positive and negative is based on Nakamura's Japanese 10 type emotion classifications [18]. Proposed emotion types projected on Russel's 2-dimensional model of affect [19] are showed on Fig. 2. The main assumption of this idea is that all emotions can be described in a space of two-dimensions: the emotions' polarity (positive/negative) and activation (activated/deactivated). The polarity (or valence) dimension refers to the hedonic tone of emotions (i.e. whether the particular emotion is pleasant or unpleasant). Activation refers to a sense of mobilization or energy, and, according to Feldman-Barrett and Russell [20], it follows the continuum ranging from sleep (low), through drowsiness, relaxation, alertness, hyperactivation, and, finally, frenetic excitement (high).

Thus, every emotive expression can be described using these two dimensions. For example, the emotive expression *wa-i wa-i!* ("hurray!") belongs to the Nakamura's group *yorokobi* (joy), and can be described as positive and active, whereas the emotive expression *shiku shiku naita yo* ("I wept!") belongs to the group *aware* (sorrow) and can be described as negative and deactivated.

In Fig. 2, some types were placed in two quarters, as they can contain both positive and negative or activated and deactivated expressions. This, however, only concerns classes of emotions — each emotive expression belongs to only one group.

4. Evaluation Experiments

After constructing the joking chatterbot (see Sect. 3.3), we conducted a number of experiments in order to evaluate its performance. The results in this paper are based the users' evaluation experiment (see Sect. 4.1) and automatic emotiveness analysis (Sect. 4.2).

4.1 Users' Evaluation

In the first experiment, we asked 13 university students to perform a conversation with the Modalin (nonhumorous) and Pundalin (humorous) systems. As the dialogue was supposed to be as free as possible, no topic restrictions were made. Users filled in a questionnaire containing questions about the dialogue. In this paper we present only results for the questions directly related to the topic:

1) Do you want to continue the dialogue? (answers in 5-point scale);

2) Which system do you think was better?

The statistical significance of results was checked using the Two Paired Sample Wilcoxon Signed Rank Test (as the data was paired, but did not have a normal distribution).

Results

11 out of 13 evaluators (85.0%) assessed Pundalin (with humor) as better than Modalin (without humor). This shows that implementation of humor influenced the general perception of conversational system. Also for the question concerning the willingness to continue the dialogue, the results point to the humor-equipped system (2.62 vs. 3.68 in 5-point scale). The results were found to be statistically significant at 5% level (one sided p value = 0.033, z score = −1.836).

4.2 Automatic Evaluation — Emotiveness Analysis

The chat logs from users' experiment were analyzed with the ML-Ask System (Sect. 3.4). Results of the analysis allowed us to compare the dialogues of our two systems (with and without humor) in three dimensions:

1) general emotiveness of conversations
A percentage of emotive utterances per 10 turns of conversation. If, for example, 6 of 10 user's utterances were found emotive, the general emotiveness of the conversation was 60.0% — as shown in Fig. 3.

2) their positivity/negativity;
If emotions were positive or changed from negative through neutral to positive over the conversation, general sentiment is considered positive. If emotions were negative or changing from positive through neutral to negative over the conversation, the general sentiment towards is considered as negative.

3) their activation/deactivation degree.
If emotional expression detected by ML-Ask belonged to the activated dimension, it was counted as activated.

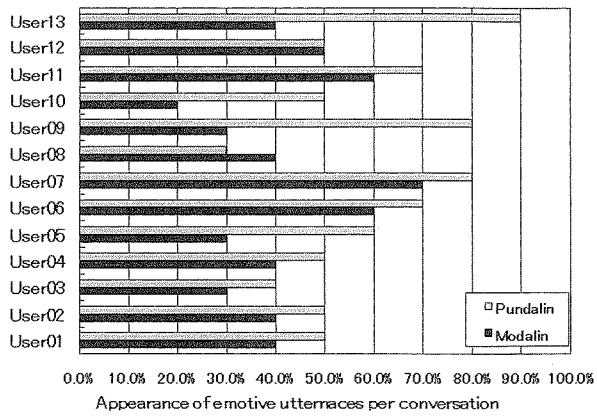


Fig. 3 Percentage of average appearance of emotively engaged utterances for all human evaluators in conversations with Modalin (non-humorous) and Pundalin (humorous system) [21].

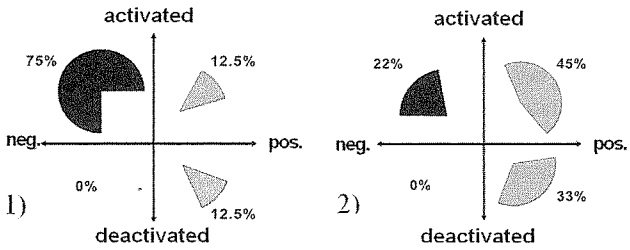


Fig. 4 Projection of emotive analysis of users' emotions types on Russell's two-dimensional space — 1) Modalin (without humor) and 2) Pundalin (with humor).

If emotional expression belonged to the deactivated dimension, it was counted as deactivated.

Results

The tendencies shown by the results of emotive analysis were generally consistent with those gained in human evaluation experiment. Almost all evaluators showed more emotions towards Pundalin — as shown in Fig. 3. The results were found to be statistically significant (the Wilcoxon test, z score = −2.706, p value = 0.003, one-sided).

While 75.0% of emotions towards Modalin were negative and activated, for Pundalin the proportions were opposite (45.0% of positive and activated, 78.0% of positive emotions in total) — see Fig. 4. In this experiment, any negative deactivated emotions were found neither in humor- nor non-humor-equipped system's chat logs.

These results are discussed in Sect. 6.

As showed on Fig. 4, most emotions towards Modalin, were negative and active (75.0%), while the proportions were opposite towards Pundalin (45.0% of emotions were positive and active, with only 22.0% of negative and active).

We have also calculated the statistical correlation between a user's willingness to continue the dialogue and the general emotiveness results showed on Fig. 3, using the Pearson product-moment correlation coefficient. The result showed that there is a rather weak, but still correlation of 0.224 between these two sets of results.

5. Engagement - Discussion

In this paper we focus on correlations between results of the experiments described above and their impact on engagement in the conversation.

5.1 Engagement and Will to Continue the Dialogue

Results of the user evaluation experiment showed that the system with humor was perceived as generally better.

However, we assume that it the question "Do you want to continue the dialogue?" as more relevant to the subject of engagement in the conversation. The statistical correlation test (see Sect. 4.2) showed a weak correlation (0.224) between the willingness to continue the dialogue and general emotiveness of users utterances. Obviously, the degree of correlation could be better, however, it gives us some information about the results:

1) There is a weak (but still) correlation between emotiveness and willingness to continue the interaction;

2) the small-scale experiment (see Sect. 1.3) showed that there is a tendency that emotiveness of dialogue and engagement of its participants are closely related — thus, we can assume that the willingness to continue the interaction is only a part of bigger feature called "engagement". However, providing this to be true, here we face another question: what are the other factors that influence engagement? This issue still needs more specification and we intend to investigate it in the future.

5.2 Engagement and Emotiveness

As indicated on Fig. 3, in the evaluation experiment almost all users showed more emotion towards Pundalin, which, based on the idea of affect-as-information and the results from our experiment (see Sect. 1.3), means that they were more emotionally engaged, when the system used humor. Therefore, it can be stated that the presence of humor generally enhances the degree of user engagement in the conversation.

However, emotiveness itself does not tell us much about the nature of engagement, which, especially in case of HCI, is of high importance. The very fact that user was emotively more engaged in the interaction does not necessarily mean that he/she liked it, as the engagement might as well be of negative nature. If, for example, a system would make user angry, emotiveness arises, but in a negative sense. Taking this into consideration, we propose to take two more aspects of emotiveness into consideration: positivity/negativity and activation/deactivation (see Sect. 5.3 and Sect. 5.4).

5.3 Positive and Negative Engagement

As shown on Fig. 4, the proportions between types of emotions conveyed by the users towards the two systems are

different. In the case of Modalin (without humor), most of emotions were negative and active. It can be thusly said were generally negatively engaged in the conversation, which is not a desirable situation. Conversely, most of the emotions shown towards Pundalin (with humor) were positive and active, while only 22.0% of them were negative and active. This leads to a conclusion that the participants were more positively engaged in the conversation with humor.

5.4 Activated/Deactivated Dimension

Here, however, we face a question: how should we treat the activation/deactivation dimension? How is it related to the positivity/negativity of emotions? Which emotional state is better for users of conversational systems — negative and active or positive and passive? Of course, the answer is quite complex and depends on the type of the system. We can imagine a situation, in which a system aimed at aiding in a user's anger management therapy, would tend to elicit and activate as many negative emotions as possible, in order to help him/her get rid of them. In such cases, active/negative emotions would probably be more desirable than passive.

Our research, however, focuses on chatterbots and the role of humor in their conversation with humans. For example, when implemented into a car navigator, it is desirable for the system to activate the driver in a positive way. Thus, it is the activated/positive quarter we should be interested in and, as shown on Fig. 4, in this aspect the difference between Modalin and Pundalin is clearly visible (12.5% vs. 45.0%).

5.5 Engagement and Humor

Basing on above conclusions, it can be said that:
1) There is a relation between the presence of humor, emotiveness of utterances and engagement in the dialogue.
2) Users were **more positively** and **less negatively** engaged in the conversation **with humor present**
3) Users were **more negatively** and **less positively** engaged in the conversation **without humor present**
4) The presence of **humor enhances** users' **positive involvement** in the conversation
5) The results are convergent with the results of the user experiment concerning their intentions to continue the dialogue.

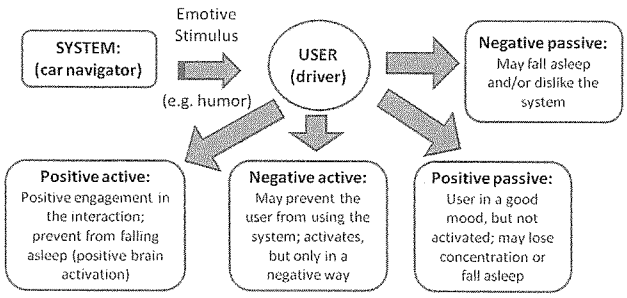


Fig. 5 Relation between emotions and types of engagement — summary (on the example of car navigator).

The summary of the results introduced in this paper is shown on Fig. 5.

6. Conclusion and Future Works

In this paper we presented results of our investigation on the correlation between the presence of humor and user involvement in the conversation with a non-task oriented conversational system. We have shown that humor can enhance positive and reduce negative engagement. This leads to the conclusion that the effect of humor is not to be neglected in the field of HCI.

The results described above can also be seen as a contribution to such fields as psycholinguistics or psychology, as, to our knowledge, no earlier works directly investigated the connection between humor, emotions, and engagement in the interaction. In the humor-equipped system presented in this paper we applied a simple "every-third-turn" rule for joke timing, which needs to be revised. Currently we are working on a timing algorithm that will recognize user emotional states and decide when it is appropriate to tell a pun. The research on this subject is still ongoing and we expect to be able to present the first results very soon.

As far as the research on correlation between engagement and emotions is concerned, the next step will be to check what specified types of emotions influence engagement in what way. In this paper we explored three dimensions of emotive states: emotiveness, positivity/negativity and activation/deactivation. Although it is a good point to start, the distinction may be too general and should be more focused.

Also the topic of humor relating to human engagement requires more research. Presumably there is a correlation between the types of humor and types of engagement, (e.g. aggressive jokes can elicit negative engagement) and this area still needs to be explored.

In this research, we focus on puns as they are relatively easy to compute using NLP methods. However, we believe that most mechanisms and discoveries (such as the correlation between engagement and humorous stimuli) should work well for other types of humor as well.

Generally, much effort is still needed to explore the areas tackled in this paper. We will be happy if it gives a new impetus to new research projects, especially concerning user engagement in the interaction with conversational computer systems.

References

[1] P. Dybala, M. Ptaszynski, S. Higuchi, R. Rzepka, and K. Araki, "Humoroids — Conversational agents that induce positive emotions with humor," Proc. AAMAS 09, pp.1171–1172, Budapest, Hungary, 2009.

[2] M. Ptaszynski, P. Dybala, S. Higuchi, R. Rzepka, and K. Araki, "How to find love in the Internet? Applying Web mining to affect recognition from textual input," Proc. EMALP'08 (PRICAI'08), pp.67–79, Hanoi, Vietnam, 2008.

[3] B. Reeves and C. Nass, The media equation: How people treat computers, television, and new media like real people and places,

Cambridge Univ., UK, 1996.

[4] C.L. Sidner and C. Lee, "Engagement rules for human-robot collaborative interactions," Proc. 2003 Conf. on Systems, Man and Cybernetics, vol.4, pp.3957–3962, Washington, USA, 2003.

[5] M. Selting, "Emphatic speech style — with special focus on the prosodic signaling of heightened emotive involvement in conversation," J. Pragmatics, vol.22, no.3-4, pp.375–408, 1994.

[6] L. Hall, S. Woods, R. Aylett, L. Newall, and A. Paiva, "Achieving empathic engagement through affective interaction with synthetic characters," ACII 05, Beijing, China, 2005.

[7] C. Yu, P.M. Aoki, and A. Woodruff, "Detecting user engagement in everyday conversations," Proc. ICSLP 2004, pp.1329–1332, Jeju Island, Korea, 2004.

[8] M. Usami, ed., BTS ni yoru nihongo hanashikotoba kōpasu 1 (hatsu taimen, yuujin; zatsudan, touron, sasoi) [Conversation corpus of spoken Japanese using the Basic Transcription System (first meeting, friend's conversation, small talk, discussion, invitation)], Tokyo University of Foreign Studies, Tokyo, Japan, 2007.

[9] N. Schwarz and G.L. Clore, "Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states," Journal of Personality and Social Psychology, vol.45, pp.513–523, 1983.

[10] K. Binsted, Machine humour: An implemented model of puns, University of Edinburgh, UK, 1996.

[11] D. Loehr, "An integration of a pun generator with a natural language robot," Proc. International Workshop on Computational Humor, eds. J. Hulstijn and A. Nijholt, pp.161–172, University of Twente, Netherlands, 1996.

[12] J. Morkes, H.K. Kernal, and C. Nass, "Effects of humor in task-oriented human-computer interaction and computer-mediated communication: A direct test of srct theory," Human-Computer Interaction, vol.14, no.4, pp.395–435, 1999.

[13] D. Mobbs, M.D. Greicius, E. Abdel-Azim, V. Menon, and A.L. Reiss, "Humor modulates the mesolimbic reward centers," Neuron, vol.40, pp.1041–1048, 2003.

[14] S. Higuchi, R. Rzepka, and K. Araki, "A casual conversation system using modality and word associations retrieved from the Web," Proc. EMNLP '08, pp.382–390, Honolulu, USA, Oct. 2008.

[15] K. Yoshihira, Y. Takeda, and S. Sekine, "KWIC system for Web Documents," Proc. 10th Annual Meetings of the Japanese Association for NLP, pp.137–139, Japan, 2004.

[16] M. Ptaszynski, Boisterous language. Analysis of structures and semiotic functions of emotive expressions in conversation on Japanese Internet bulletin board forum-2channel, M.A. Dissertation, UAM, Poznan, Poland, 2006.

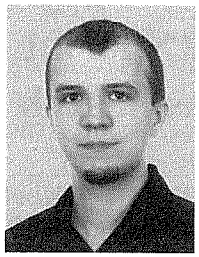
[17] T. Sugiyama Lebra, The Japanese Self in Cultural Logic, University of Hawaii Press, Hawaii, USA, 2004.

[18] A. Nakamura, Kanjo hyogen jiten [Dictionary of Emotive Expressions], Tokyodo Publishing, Tokyo, Japan, 1993.

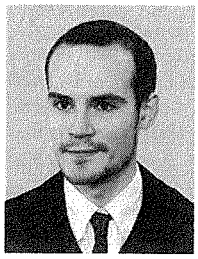
[19] J.A. Russell, "A circumplex model of affect," Journal of Personality and Social Psychology, vol.39, no.6, pp.1161–1178, 1980.

[20] L. Feldman Barrett and J.A. Russell, "The structure of current affect: Controversies and emerging consensus," Current Directions in Psychological Science, vol.8, no.1, pp.10–14, 1999.

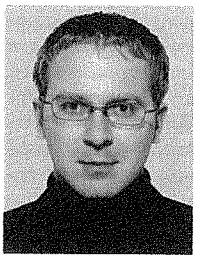
[21] M. Ptaszynski, P. Dybala, S. Higuchi, R. Rzepka, and K. Araki, "Affect as information about users' attitudes to conversational agents," Proc. HAI'08 (WI-IAT'08), pp.459–500, Sydney, Australia, 2008.



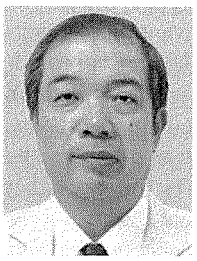
Pawel Dybala was born in Ostrow Wielkopolski, Poland in 1981. He received his M.A. from the Jagiellonian University in Krakow, Poland in 2006. He was a research student at Hokkaido University, and since 2007 he is studying towards his Ph.D. degree at the Graduate School of Information Science and Technology, Hokkaido University, Japan. His research interests include NLP, Dialogue Processing, Humor Processing, HCI and Information Retrieval.



Michal Ptaszynski was born in Wroclaw, Poland in 1981. He received his M.A. from the University of Adam Mickiewicz in Poznan, Poland in 2006. He was a research student at Otaru University of Commerce, and since 2007 he is studying towards his Ph.D. degree at the Graduate School of Information Science and Technology, Hokkaido University, Japan. His research interests include NLP, Dialogue Processing, Affect Analysis, Sentiment Analysis, HCI and Information Retrieval. He is a member of the SOFT, JSAI and NLP.



Rafal Rzepka was born in Szczecin, Poland in 1974. He received his M.A. from the University of Adam Mickiewicz in Poznan, Poland in 1999 and Ph.D. from Hokkaido University, Japan in 2004. Now he is an assistant professor at the Graduate School of Information Science and Technology, Hokkaido University, Japan. His research interests include NLP, Web Mining, Commonsense Retrieval, Dialogue Processing, Language Acquisition, Affect and Sentiment Analysis. He is a member of the AAAI, ACL, JSAI, IPSJ, CSS and NLP.



Kenji Araki was born in 1959 in Otaru, Japan. He received B.E., M.E. and Ph.D. degrees in electronics engineering from Hokkaido University, Sapporo, Japan in 1982, 1985 and 1988, respectively. In April 1988, he joined Hokkai Gakuen University, Sapporo, Japan. He was a professor of Hokkai Gakuen University. He joined Hokkaido University in 1998 as an associate professor of the Division of Electronics and Information Engineering. He was a professor of the Division of Electronics and Information Engineering of Hokkaido University from 2002. Now he is a professor of the Division of Media and Network Technologies of Hokkaido University. His research interests include NLP, Spoken Dialogue Processing, Machine Translation and Language Acquisition. He is a member of the AAAI, IEEE, JSAI, IPSJ, and JCSS.