

Natural Language Dialog System Considering Speaker's Emotion Calculated from Acoustic Features

Takumi Takahashi¹, Kazuya Mera¹, Tang Ba Nhat², Yoshiaki Kurosawa¹, and
Toshiyuki Takezawa¹

¹ Graduate School of Information Sciences, Hiroshima City University, 3-4-1, Ozuka-higashi,
Asa-minami-ku, Hiroshima 731-3194, Japan

{takahashi, tang}@is.info.hiroshima-cu.ac.jp

{mera, kurosawa, takezawa}@hiroshima-cu.ac.jp

² FPT Software, Japan

Abstract. With the development of Interactive Voice Response (IVR) systems, people can not only operate computer systems through task-oriented conversation but also enjoy non-task-oriented conversation with the computer. When an IVR system generates a response, it usually refers to just verbal information of the user's utterance. However, when a person gloomily says "I'm fine," people will respond not by saying "That's wonderful" but "Really?" or "Are you OK?" because we can consider both verbal and non-verbal information such as tone of voice, facial expressions, gestures, and so on. In this paper, we propose an intelligent IVR system that considers not only verbal but also non-verbal information. To estimate a speaker's emotion (positive, negative, or neutral), 384 acoustic features extracted from the speaker's utterance are utilized to machine learning (SVM). Artificial Intelligence Markup Language (AIML)-based response generating rules are expanded to be able to consider the speaker's emotion. As a result of the experiment, subjects felt that the proposed dialog system was more likable, enjoyable, and did not give machine-like reactions.

Keywords: Interactive Voice Response system (IVR) · Acoustic features · Emotion · Support Vector Machine (SVM) · Artificial Intelligence Markup Language (AIML)

1 Introduction

Interactive Voice Response (IVR) systems are rapidly spreading in our daily life, e.g. Siri by Apple Inc. and some advanced car navigation systems. Although previous IVR systems can only deal with task-oriented dialog that responds to the user's demands such as answering a user's question, executing a user's order, and recommending useful items, recent IVR systems attempt non-task-oriented dialog like chat and small talk. These kinds of functions are useful for people to not only waste time but also increase their affinity with and tolerance of computer systems.

When standard IVR system generates a response to the speaker's utterance, only verbal information of the utterance is considered. However, when a person gloomily says "I'm fine," people will respond not by saying "That's wonderful" but "Really?" or

“Are you OK?” because we can consider both verbal and non-verbal information such as tone of voice, facial expressions, gestures, and so on.

In this paper, we propose an intelligent IVR system that considers not only verbal but also non-verbal information. To estimate a speaker’s emotion (positive, negative, or neutral), 384 acoustic features extracted from the speaker’s utterance are utilized to machine learning (SVM). Artificial Intelligence Markup Language (AIML) based response generating rules are expanded to be able to consider both text expressions and the speaker’s emotion as condition parts of the rules.

Section 2 introduces the overview of our proposed method. Sections 3 and 4 explain an expansion of AIML-based response generating rules and an emotion estimating method from acoustic features, respectively. Section 5 discusses experimental results. The conclusion and future works are presented in Section 6.

2 Overview of the Proposed Method

The proposed method is based on a stimulus-response style response generating system. Response generating rules usually consist of a condition part that matches input utterance and a response part that generates response sentence. The general rule-based method compares the verbal expression of input utterance with the condition part of the rules. However, even if a speaker utters completely the same sentence, utterances with different non-verbal expressions often make different impressions as discussed in Section 1.

In this paper, we propose a method that can consider both verbal expressions and the speaker’s emotion estimated from acoustic features when the method selects the most adequate response generating rule. Figure 1 shows the overview of our proposed method. To carry out pattern matching process, two types of information (i.e. verbal and non-verbal) are required. As verbal information, the text expression of the speaker’s utterance is used. It is obtained by a speech recognition unit, and the text is analyzed morphologically because Japanese sentences are written without spaces between words.

On the other hand, the speaker’s emotion is estimated from acoustic features of the utterance. First, acoustic features of the utterance are calculated. Next, the speaker’s emotion is estimated from the acoustic features by using a machine learning unit that previously learned the relationship between emotion types and the acoustic features. The pattern matching process compares the condition part of all rules with the speaker’s utterance. Condition of the rule specifies both text features and emotion state. Finally, response sentence is generated on the basis of the response part of the selected rule.

Section 3 introduces the expansion of AIML-based response generating rules, and Section 4 describes the emotion estimating method from acoustic features.

3 Expansion of AIML-based Response Generating Rule

There are response generating methods based on the stimulus-response model. One of the most famous models is Artificial Intelligence Markup Language (AIML), which is an XML-compliant language [1]. Artificial Linguistic Internet Computer Entity (A.L.I.C.E),

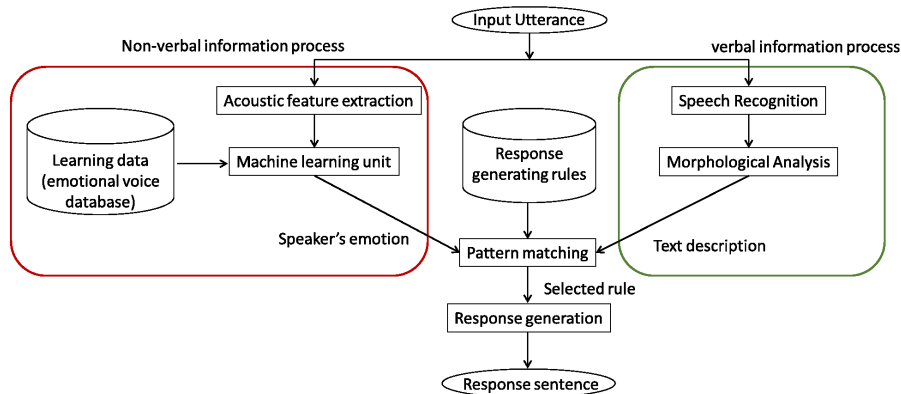


Fig. 1. Overview of our proposed method

which has won the annual Loebner Prize Competition in Artificial Intelligence [2] three times, is also constructed by using AIML.

The most important units of AIML are the follows:

- *< aiml >*: the tag that begins and ends an AIML document
- *< category >*: the tag that marks a “unit of knowledge” in an Alicebot’s knowledge base
- *< pattern >*: used to contain a simple pattern that matches what a user may say or type to an Alicebot
- *< template >*: contains the response to a user input

The AIML pattern language is consisting only of words, spaces, and the wildcard symbols *_* and ***. Template language is described by not only complete sentences but also “template” expressions that are filled up with additional words inherited from *< pattern >* or other units.

Our proposed method improves *< pattern >* unit of AIML in order to consider both text information and the speaker’s emotion. The *< pattern >* part of our proposed method consists of three conditions: “Text,” “Previous utterance,” and “Emotion.” “Text” is the text of a speaker’s utterance. It corresponds to the *< pattern >* unit of AIML. “Previous utterance” refers to the robot’s previous utterance. In AIML, the syntax *< that > ... < /that >* encloses a pattern that matches the robot’s previous utterance. “Emotion” is our original factor that indicates the speaker’s emotion estimated from acoustic features of the input utterance. Figure 2 shows an example of the response rule matching process. When a speaker cheerfully utters “Good morning,” the speaker’s emotion is estimated to be “happy,” and rule A is selected. As a result, the system responds with “Good morning.” On the other hand, when the speaker gloomily utters “Good morning,” the speaker’s emotion is estimated “sad,” and rule B is selected. As a result, the system responds “You don’t sound happy. What’s up?” As shown in the example, our proposed method can respond with an appropriate reaction by considering the speaker’s emotion.

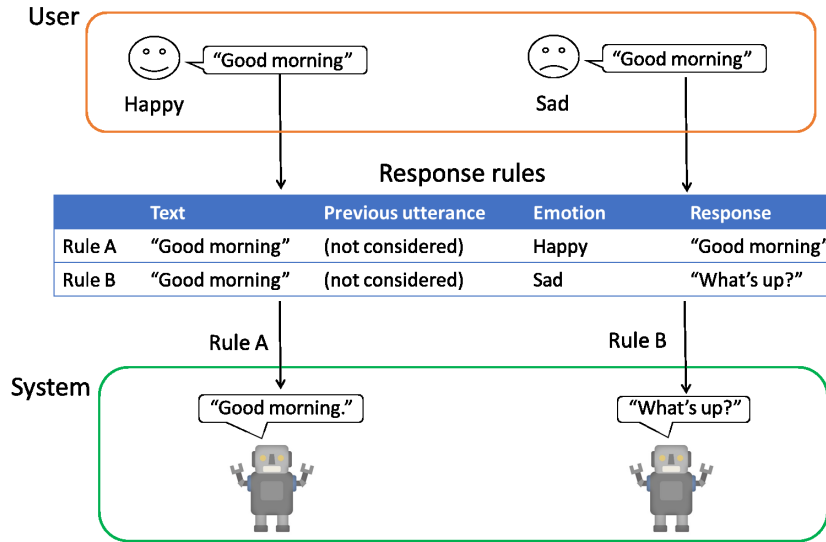


Fig. 2. Example of Response Generation

4 Emotion Estimation Method based on Acoustic Features

The young field of emotion recognition from voices has recently gained considerable interest in human-machine communication, human-robot communication, and multimedia retrieval. Numerous studies have been seen in the last decade trying to improve on features and classifiers. Therefore, the "Emotion Challenge" competition, in which various emotion recognition methods competed, was held at INTERSPEECH 2009 [3]. In the competition, open source software (openSMILE [4]) was provided that can extract various features. Schuller exhibited baseline results for 2-class emotion recognition, NEGative (subsuming *angry*, *touchy*, *reprimanding*, and *emphatic*) and IDLe (consisting of all non-negative states), by using 384 features. As a result, weighted averages of the recall and precision were 0.73 and 0.71, respectively.

Our proposed method also uses the feature set used at the INTERSPEECH 2009 Emotion Challenge that calculates 384 static feature values. The feature set refers to the following 16 low-level descriptors (contours):

- **pcm_RMSenergy**: Root-mean-square frame energy
- **mfcc**: Mel-Frequency cepstral coefficients 1-12
- **pcm_zcr**: Zero-crossing rate of time signal (frame-based)
- **voiceProb**: The voicing probability computed from the ACF (Autocorrelation Function)
- **F0**: The fundamental frequency computed from the Cepstrum

Because a 1st order delta coefficient (differential) of each contour is calculated for each contour, the number of contours becomes 32. The following 12 features are calculated for each contour:

- **max**: The maximum value of the contour
- **min**: The minimum value of the contour
- **range** = max – min
- **maxPos**: The absolute position of the maximum value (in frames)
- **minPos**: The absolute position of the minimum value (in frames)
- **amean**: The arithmetic mean of the contour
- **linregc1**: The slope (m) of a linear approximation of the contour
- **linregc2**: The offset (t) of a linear approximation of the contour
- **linregerrQ**: The quadratic error computed as the difference in the linear approximation and the actual contour
- **stddev**: The standard deviation of the values in the contour
- **skewness**: The skewness (3rd order moment)
- **kurtosis**: The kurtosis (4th order moment)

Our proposed method uses a Support Vector Machine (SVM) as a machine learning unit. For machine learning, 26 positive voices, 29 negative voices, and 30 neutral voices were prepared. They were collected from TV dramas and animations and were annotated with one type of emotion (positive, negative, and neutral) by five university students. Only the data in which all annotators annotated the same emotion type were used.

The performance of the trained classifier was measured by leave-one-out cross validation. Table 1 is a confusion matrix of the emotion estimation result. Although accuracy and precision of “neutral” were good, “positive” and “negative” sometimes confused. The total accuracy was 0.71. Although two types of pre-processing strategies (standardization after balancing by SMOTE and balancing after standardization) were also applied, both accuracies (0.71) were the same as that of the non-pre-processing strategies.

Table 1. Confusion matrix of emotion estimation result

| | | predict | | | |
|--------|-----------|----------|---------|----------|--------|
| | | positive | neutral | negative | recall |
| actual | positive | 15 | 0 | 11 | 0.58 |
| | neutral | 0 | 27 | 3 | 0.90 |
| | negative | 11 | 0 | 18 | 0.62 |
| | precision | 0.58 | 1.00 | 0.56 | |

5 Experiment

A dialog system that can consider speaker’s emotion on the basis of the proposed method was constructed to evaluate the impression for the dialog agent. Julius[5] and openSMILE[4] were utilized for speech recognition and acoustic feature extraction,

respectively. SVM in a statistical computing environment R[6] was utilized for machine learning. Conversation logs by the previous method, which considers text expression only, and by the proposed method, which can consider both text expression and speaker’s emotion, were evaluated from the point of view of likeability, machine-creature likeness, and feeling of using. Table 2 contains two conversation logs by the previous method. Tables 3 and 4 are conversation logs that proposed method estimated speaker’s emotion correctly and inadequately, respectively. Table 5 is a conversation log that proposed method succeeded and failed to estimate speaker’s emotion in a conversation.

Table 2. Conversation logs by previous method

| | |
|---------|-------------------------------------|
| User: | Hello. |
| System: | Hello. Do you have plans for today? |
| User: | Not much. (<i>Said happily</i>) |
| System: | I understand. |
| User: | Will it be fine tomorrow? |
| System: | It will rain. |
| User: | Rain... (<i>Said gloomily</i>) |
| System: | You should take an umbrella. |

Table 3. Conversation logs by proposed method (correct estimation)

| | |
|---------|--|
| User: | Hello. |
| System: | Hello. Do you have plans for today? |
| User: | Not much. (<i>Said happily</i>) |
| System: | Really? Your voice is bouncy! |
| User: | Will it be fine tomorrow? |
| System: | It will rain. |
| User: | Rain... (<i>Said gloomily</i>) |
| System: | Is it bad for you? |
| User: | I wanted to go on a picnic tomorrow. |
| System: | It’s so pity. But, the weather will get better from afternoon. |

In order to simulate these dialog examples, 20 improved AIML rules as shown in Table 6 were added into our proposed dialog system. Format of the rules is explained in Section 3.

All of the conversation scenes were presented in videos to subjects (nine males and a female; 22-24 years old). All of the videos were about a minute long. The subjects gave their impressions for each conversation scene after they finished watching each video.

In this experiment, the conversations were evaluated by a “likeability factor” with 15 adjectives, “personality factor” with four adjectives, and “machine-creature likeness”

Table 4. Conversation logs by proposed method (inadequate estimation)

| | |
|---------|-------------------------------------|
| User: | Hello. |
| System: | Hello. Do you have plans for today? |
| User: | Not much. (<i>Said gloomily</i>) |
| System: | Really? Your voice is bouncy! |
| User: | Will it be fine tomorrow? |
| System: | It will rain. |
| User: | Rain... (<i>Said happily</i>) |
| System: | Is it bad for you? |
| User: | Not at all. |

Table 5. A conversation log by proposed method (including both correct and inadequate estimations)

| | |
|---------|--|
| User: | Please recommend a movie. |
| System: | How about "Friday the 13th?" |
| User: | What kind of movie is it ? |
| System: | It's a horror movie. |
| User: | Hmm, Horror movie... (<i>Said gloomily</i>) |
| System: | Don't you like horror movies? Then, how about "Stand by me?" |
| User: | Hmm, Stand by me... (<i>Said acceptably</i>) |
| System: | Don't you like "Stand by me," too? |
| User: | Yes, I like it. Thank you. (<i>Said happily</i>) |
| System: | You're welcome. |

Table 6. Added AIML rules especially for this experiment

| Text | < pattern > | | Response |
|--------------------------------------|-------------------------------------|----------|--|
| | Previous utterance | Emotion | |
| Hello. | * | * | Hello. Do you have plans for today? |
| Not much. | Hello. Do you have plans for today? | positive | Really? Your voice is bouncy! |
| Not much. | Hello. Do you have plans for today? | neutral | I understand. |
| Not much. | Hello. Do you have plans for today? | negative | May I recommend something amusing? |
| Will it be fine tomorrow? | * | * | It will rain. |
| Rain. | It will rain. | positive | Is it happy for you? |
| Rain. | It will rain. | neutral | You should take an umbrella. |
| Rain. | It will rain. | negative | Is it bad for you? |
| I wanted to go on a picnic tomorrow. | Is it bad for you? | * | It's so pity. But, the weather will get better from afternoon. |
| Please recommend a movie. | * | * | How about "Friday the 13th?" |
| What kind of movie is it? | How about "Friday the 13th?" | * | It's a horror movie. |
| Horror movie. | It's a horror movie. | positive | Yes! Do you like it? |
| Horror movie. | It's a horror movie. | neutral | Yes. |
| Horror movie. | It's a horror movie. | negative | Don't you like horror movies? Then, how about "Stand by me?" |
| Stand by me. | How about "Stand by me?" | positive | Yes! Do you like it? |
| Stand by me. | How about "Stand by me?" | neutral | Yes. |
| Stand by me. | How about "Stand by me?" | negative | Don't you like "Stand by me," too? |
| Thank you. | * | positive | You're welcome |
| Thank you. | * | neutral | You're welcome. |
| Thank you. | * | negative | Were you really satisfied? |

with two adjectives as shown in Table 7. These factors were selected with reference to the work Takayoshi and Tanaka [7], who evaluated impressions of a robot. Furthermore, “feeling of using” was also evaluated by using three adjectives.

Table 7. Evaluation factors for computer dialog system

| Factors | Adjectives |
|---------------------------|---|
| Likeability factor | diplomatic, want to mimic, cool, pleasing, want to be friends, want to see again, gentle, arouse sympathy, sensible |
| Personality factor | faithfully, calm, mischievous, likely unfaithfully |
| Machine-creature likeness | machine-like, creature-like |
| Feeling of using | enjoyable, annoying, systematic |

The subjects evaluated all conversations from the point of view of each adjective by a five-grade evaluation (+2:think so, +1:slightly think so, 0:neither, -1:do not much think so, -2:do not think so).

The mean value and the standard deviation for each adjective are shown in Table 8. In addition, the list of significant different adjectives are shown in Table 9, 10, 11, 12, and 13.

About likeability factor: Likeability of a dialog system directly relates to being personable. The proposed method that could estimate the speaker’s emotions correctly had significantly higher mean values of adjectives about likeability than the previous method, especially “pleasing ($p < 0.001$)” and “want to be friends ($p < 0.001$).” On the other hand, the proposed method that estimated inadequate emotions had significantly lower mean values of “want to mimic ($p < 0.05$)” and “cool ($p < 0.01$)” than the previous method.

About machine-creature likeness factor: The result denoted that the subjects felt that the previous method was more “machine-like ($p < 0.001$)” and the proposed method was more “creature-like ($p < 0.001$)” even if the proposed method estimated inadequate emotion (machine-like: $p < 0.001$, creature-like: $p < 0.001$). Although the previous method always responded with the same response to the same text expression, the proposed method changed responses in accordance with the speaker’s emotion. This provides two impressions: the previous method is monotonous, and the proposed method is not machine-like because it can deal with emotion.

About feeling of using factor: The proposed method that could estimate the speaker’s emotions correctly had a significantly higher mean value of “enjoyable ($p < 0.001$)” than the previous method. On the other hand, the previous method had a significantly higher mean value of “systematic ($p < 0.001$)” than the proposed method. Subjects could enjoy the reaction of the proposed method because changing response patterns in accordance with acoustic features of utterance was novel. However, the proposed method that estimated inadequate emotions had a significantly higher mean value of “be annoying ($p < 0.001$)” than that in succeed situation.

Table 8. Mean value and standard deviation for each factor and adjective

| | | | Previous method (Table 2) | Proposed method | | |
|---------------------------|--------------------|-------|------------------------------|---------------------------------|------------------------------------|-----------------------------------|
| | | | | Correct estimation (Table 3) | Inadequate estimation (Table 4) | Imperfect estimation (Table 5) |
| | | N | 10 | 10 | 10 | 10 |
| Likeability factor | diplomatic | Mean | 0.35 | 1.60 | 0.20 | 1.90 |
| | | SD | 1.15 | 0.58 | 1.12 | 0.30 |
| | want to mimic | Mean | 0.05 | 0.70 | -0.50 | 1.00 |
| | | SD | 0.86 | 0.78 | 0.59 | 0.89 |
| | cool | Mean | 0.10 | 0.20 | -0.50 | 0.80 |
| | | SD | 0.70 | 0.68 | 0.50 | 0.87 |
| | pleasing | Mean | -0.05 | 1.35 | -0.5 | 1.30 |
| | | SD | 1.12 | 0.96 | 1.28 | 1.19 |
| | want to be friends | Mean | -0.35 | 0.95 | -0.30 | 1.20 |
| | | SD | 1.24 | 1.07 | 1.05 | 1.17 |
| | want to see again | Mean | -0.40 | 0.85 | -0.40 | 1.20 |
| | | SD | 0.97 | 1.11 | 0.80 | 1.17 |
| | gentle | Mean | 0.20 | 1.35 | 0.00 | 1.70 |
| | | SD | 1.08 | 0.65 | 0.89 | 0.46 |
| arouse sympathy | Mean | -0.05 | 0.95 | -0.45 | 1.00 | |
| | SD | 0.97 | 0.97 | 0.80 | 1.10 | |
| sensible | Mean | 0.10 | 1.05 | -0.55 | 1.90 | |
| | SD | 1.48 | 1.24 | 1.36 | 0.30 | |
| Personality factor | faithfully | Mean | 1.05 | 0.40 | 0.25 | 0.40 |
| | | SD | 0.74 | 0.80 | 0.99 | 0.80 |
| | calm | Mean | 0.40 | -0.55 | 0.10 | -0.20 |
| | | SD | 0.80 | 0.74 | 0.94 | 0.60 |
| | mischievous | Mean | -0.80 | -0.20 | -0.20 | -0.10 |
| | | SD | 0.93 | 0.93 | 0.93 | 0.54 |
| | unfaithfully | Mean | -0.75 | -0.80 | -0.30 | -1.00 |
| | | SD | 0.83 | 0.60 | 1.10 | 0.63 |
| Machine-creature likeness | machine-like | Mean | 1.35 | -1.15 | 0.50 | -1.00 |
| | | SD | 0.79 | 0.91 | 0.97 | 1.10 |
| | creature-like | Mean | -0.90 | 1.30 | -0.15 | 1.20 |
| | | SD | -0.89 | 0.64 | 0.96 | 0.98 |
| Feeling of using | enjoyable | Mean | -0.50 | 1.20 | -0.40 | 1.30 |
| | | SD | 0.87 | 0.40 | 0.80 | 0.64 |
| | be annoying | Mean | -0.10 | -1.00 | 0.30 | -0.60 |
| | | SD | 0.83 | 1.41 | 1.00 | 1.56 |
| | systematic | Mean | 0.90 | -0.90 | 0.65 | -1.00 |
| | | SD | 0.89 | 0.77 | 0.73 | 0.89 |

N: number of subjects

Table 9. A list of significant different adjectives between previous method and proposed method (correct estimation)

| Previous method | Proposed method (correct estimation) |
|--|--|
| faithfully*, calm**, machine-likeness***, be annoying**, systematic*** | diplomatic***, want to mimic*, pleasing**, want to be friends***, want to see again***, gentle**, arouse sympathy**, sensible*, creature-likeness***, enjoyable*** |

*** $p < .001$ ** $p < .01$ * $p < .05$

Table 10. A list of significant different adjectives between previous method and proposed method (inadequate estimation)

| Previous method | Proposed method (inadequate method) |
|--|-------------------------------------|
| want to mimic*, cool**, faithfully*, machine-likeness*** | unfaithfully*, creature-likeness** |

*** $p < .001$ ** $p < .01$ * $p < .05$

Table 11. A list of significant different adjectives between previous method and proposed method (imperfect estimation)

| Previous method | Proposed method (imperfect estimation) |
|---|--|
| faithfully*, machine-likeness***, systematic*** | diplomatic***, want to mimic*, cool*, pleasing**, want to be friends**, want to see again***, gentle***, arouse sympathy*, sensible***, mischievous*, creature-likeness***, enjoyable*** |

*** $p < .001$ ** $p < .01$ * $p < .05$

Table 12. A list of significant different adjectives between proposed method (correct estimation) and proposed method (inadequate estimation)

| Proposed method (correct estimation) | Proposed method (inadequate estimation) |
|---|---|
| diplomatic***, want to mimic***, cool**, pleasing***, want to be friends***, want to see again***, gentle**, arouse sympathy**, sensible*, creature-likeness***, enjoyable*** | calm*, unfaithfully* machine-likeness***, be annoying***, systematic*** |

*** $p < .001$ ** $p < .01$ * $p < .05$

Table 13. A list of significant different adjectives between proposed method (correct estimation) and proposed method (imperfect estimation)

| Proposed method (correct estimation) | Proposed method (imperfect estimation) |
|--------------------------------------|--|
| - | sensible** |

*** $p < .001$ ** $p < .01$ * $p < .05$

Comparison between perfect and imperfect emotion estimation: The impressions between the system which estimated the speaker’s emotions perfectly and that sometimes estimated inadequate emotions were almost the same except “sensible ($p < 0.01$).” It denotes that the likeability of the dialog system does not decrease so much even if the system sometimes fail to estimate the speaker’s emotion.

6 Conclusion

In this paper, we proposed a natural language dialog method that can consider both text expressions and the speaker’s emotion estimated from acoustic features. The speaker’s emotion was estimated by using SVM based on 384 acoustic features extracted by openSMILE. Response generating rules were expanded from the AIML format. The $\langle pattern \rangle$ part of the rule consisted of three conditions: “Text,” “Previous utterance,” and “Emotion.”

The impressions of conversations by the previous method and the proposed method were evaluated from the point of view of “likeability factor,” “personality factor,” “machine-creature likeness,” and “feeling of using.” As a result, subjects felt the proposed method was more personable, creature-like, and enjoyable than previous method. Furthermore, the positive impression was kept up even if the emotion estimation result was not perfect.

For the future work, the number of response generating rules should be increased in large quantities to realize an intellectual and flexible dialog system. Furthermore, not only acoustic feature but also other non-verbal features like facial expressions should be considered to improve the accuracy of emotion estimation.

Acknowledgement

This research is supported by JSPS KAKENHI Grant Number 26330313 and the Center of Innovation Program from Japan Science and Technology Agency, JST.

References

1. AIML - The Artificial Intelligence Markup Language, <http://www.alicebot.org/aiml.html>, (Accessed May. 9, 2016)

2. Home Page of the Loebner Prize, <http://www.loebner.net/Prizef/loebner-prize.html>, (Accessed May. 9, 2016)
3. Emotion Challenge - AAAC emotion-research.net -Association for the Advancement of Affective Computing, <http://emotion-research.net/sigs/speech-sig/emotion-challenge>, (Accessed May. 9, 2016)
4. Eyben, F., Wöllmer, M., Schuller, B.: openSMILE: The Munich Versatile and Fast Open-Source Audio Feature Extractor. In: Proceedings of the International Conference on Multimedia (2010)
5. Lee, A., Kawahara, T., Shikano, K.: Real-time confidence scoring based on word posterior probability on two-pass search algorithm. Technical Report of IEICE 103(520), pp. 35–40 (2003), (in Japanese)
6. Ihaka, R., Gentleman, R.: R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics* 5, pp. 299–314 (1996)
7. Takayoshi, K., Tanaka, T.: The relationship between the behavior of kuwata and impression of his intelligence & personality. *IPSJ SIG Technical Report* 160, 43–48 (2007), (in Japanese)