ESTIMATING EMPATHIC STATES IN A SPOKEN DIALOGUE DATA BY USING SPEECH PARAMETERS

Yahan Gao, Risu Na, Yuko Ohno, Yoshihiko Hayashi
Shanghai Medical Instrumentation College
Shanghai Yinkou Road 101/ China
gaoyh@smic.edu.cn

ABSTRACT
We have been investigating characteristics of empathic states expressed in spontaneous utterances by analyzing speech data annotated with attitude and emotion categories. Based on these investigations, this paper examines a method to estimate empathic states from speech utterances. The idea is to use the annotated data as training data for a machine learning algorithm (SVM) to construct relevant classifiers. These classifiers are arranged in such a way to reflect our definition of empathic states. The experimental results suggest: (1) Dependencies exist between attitude and emotion; (2) The estimation can be improved if augmented with the categories for the preceding utterance. This paper further discusses details of the initial experimental results and proposes a research agenda for future studies.

KEY WORDS
Empathic state, Speech parameters, Machine learning Spontaneous dialogue.

1. Introduction

It is of great importance to show "empathy" to the other person involved at the right time and the right scene in order to progress a smooth dialogue. The empathy in a dialogue is expressed by facial expressions, gestures, speeches, and speech contents transmitted to the other party. The empathic status in this study is defined for the utterance Y responding to a previous utterance X, and it is expressed as a set of two pieces of information: (a) the "attitude" to the topic contents of utterance X. (b) information on status of the emotion of utterance X and Y. Among these, "empathy" is defined as the empathy state in which both the attitude and emotions are the same in particular, being different from other empathy states. Based on the annotated data obtained by annotation with the categories of emotion and attitude on the audio data of the spontaneous speech, we have analyzed voices of the utterance concerning the "empathy state", and have drawn the following conclusions: (1) The emotions and attitudes of utterance are not independent. (2) Phonetic features are more distinct in emotion than attitude, (3) In order to estimate the empathy state, and also to estimate the emotions and attitudes of every utterance, it is necessary to identify the analysis object reacted to a certain previous utterance. [Gaoyahan, 2010a; Gaoyahan, 2010b].

In this study, we examine the estimation of empathy state based on (SVM) machine learning by using annotated data as training data. The method reported in this study is to use the annotated data as training data for a machine learning algorithm (SVM) to construct relevant classifiers. These classifiers are arranged in such a way to reflect our definition of empathic states. In other words, an identification process consisting of the following steps is proposed: the first step is to identify whether the target utterance is a new topic. If a new topic is identified, it is possible to detect the "topic section" from utterances corresponding to certain topics. This is the second step. Then, based on the assumption that the reaction utterance of a target utterance has been identified by some means, identify "attitude" and "emotion".

In this paper, the definitions of empathy and empathy state are defined in Section 2, and the analysis and annotation of dialogue are presented in Section 3, then the estimation method of empathy state by machine learning is described in Section 4. And study results are detailed in Section 5. In addition, result summary and a discussion on future researches are set in Section 6.

2. The Empathy and Empathic Statuses

2.1 Empathy (a subclass of the empathic state)

The full name of authors must be written as the first name, followed by the middle initial and last name. For example, Alex B. Falcon, where "Alex" is the first name, "B." is the initial for the middle name, and "Falcon" is the last name.

Empathy is defined in many different ways. Famous definitions of empathy often fall into the following two categories: (1) cognitive empathy – "the intellectual/imaginative apprehension of another's mental state" and (2) emotional empathy – "an emotional response to …emotional responses of others" [Mehrabian 1972]. Both of them define empathy as an internal mental state. In this study, we assume that empathy is "the speaker's attitude and emotional state that are congruent with the one he or she is interacting with". This is different with the case of that attitude congruent, and

DOI: 10.2316/J.2014.216.815-0029

emotion incongruent. The former is from cognitive aspect, and the latter is from the emotional aspect. All these define empathy as an internal mental state.

From the aspect of smooth communication, we focus on the expression of attitude and emotional state related to empathetic communication. In our study, the combination of emotional states and attitudes are referred to "empathic statuses". Empathic statuses are expressed by a variety of modes, but we focus on the sound of utterance in particularly. In particular, from the aspect of smooth communication, we focus on expression of speaker's emotions and attitudes towards the person he or she is interacting with.

In other words, at least, the speaker's perception towards the topic (attitude) and his or her feelings on the topic (emotion) are considered. By expressing these properly, "empathy" or "non-empathy" of "empathic state" is communicated to the other person. Empathy is defined as a subclass of one of the "empathic states" as shown in Table1.

Table 1: The Definition of Empathic State

Attitude	Emotion	Empathic state
Positive	Congruent	Empathy
	Incongruent	Positive+ Emotion incongruent
Negative	Congruent	Negative+ Emotion incongruent
	Incongruent	Totally incongruent
Neutral	Congruent	Emotion congruent
	Incongruent	Emotion incongruent

The expression of empathic state uses a variety of modes, but we focus on the features in the speech. It is still difficult to obtain the linguistic content of speech in a free dialogue with speech recognition and language processing at present. Relatively robust voice feature is possible to be obtained. By combining with information from other modes, empathic statuses estimated from the audio information is expected to contribute to the estimation of conversational "activity field" [Maeda 2010]. Further, empathy is very important in Speech Analysis of the Empathic Statuses in Free Conversations consultation service of the hospital. Empathy plays an important role in caring for and healing the patient. It is hoped that through this study, a guideline for this can be obtained [Gaoyahan, 2009].

2.2 The Definition of "Empathic State"

The definition of "empathic state" in this study is depended on the speech segment Y that is reacted to the preceding speech segment X, and is expressed as a set of the two aspects of information: (a) "attitude" for the topic content of utterance X, (b) the match state of emotion in utterance X and Y. The definition of "empathic state" in this study is shown in Table 1 below. Both "attitude" and "emotion" are matched with "empathic state" which is called "empathy" therein. Distinguish it from another empathy state. The following is classification categories of emotions and attitudes in this study.

The "attitude" is defined for the speech segment Y react to the speech segment X of the preceding interacting, speaker of the speech segment Y have the same opinions and ideas or not to the content of the speech segment X. It is sub-classified as follows depending on whether "attitude" is the same.

- Positive: The opinions and ideas in speech segment Y is the same with content of the speech segment X.
- Negative: The opinions and ideas in speech segment Y are different with content of the speech segment X.
- Neutral: No clear opinions and ideas for content of the speech segment X are shown.

Furthermore, when presenting completely new topics, reaction target is not present, so in this time "attitude" cannot be defined, we use "Φ" to mark "attitude" in this case.

The classification of emotional state adopts basic categories for emotion which are proposed by Ekman [2], and a state of "calm" is added. [Ekman, 1992]

- The classification of Ekman's basic emotions: "joy", "anger", "sadness" and "fear", "disgust" and "surprise".
- Calm: emotions are not explicitly expressed.

3. The Unit of Analysis and Flow of the Topic

In order to analyze the dialogue data, it is necessary to determine the basic unit. In this study, the basic unit of analysis is referred to as a "speech segment", which can be determined by a time period concerning the same opinion or idea in a speech.

From the perspective of the flow of a conversation topic, each speech segment can be classified as follows. This called a "topic function" of the speech.

- The present of a new topic: The present of a new topic at the beginning of the dialogue, or the change of topic during the dialogue. And the new topic is not related to the current one in the dialogue.
- The topic expansion: the topic is changed, but it remains related to the current topic.
- The topic continuation: the current topic is maintained.

A time period with the same topic is defined as a "topic section".

4. The Unit of Analysis and Flow of the Topic

4.1 The Dialogue Data

The object data was a part of the data from a three-person interactive dialogue experiment about research of "activity field" [Maeda 2010]. The subjects in this study were college students, 6 male (age: 21.5±0.8) and 6

female (age: 21.0±1.8). It was the first time for them to meet each other, and they were grouped by gender and were asked to talk with each other. In order to smoothly start the three-person dialogue, they were asked to talk freely with pair (in advance for 2 minutes). Then the three-person dialogue started and were conducted twice, per dialogue lasted for 4 minutes. In the end, we got four sets of dialogue data (8 minutes per set, total of 32 minutes). The specific topic often exists in the conversation of first meeting. Therefore, specific topics (such as money) were given for the second dialogue. The participants performed a conversation in a free situation where a headset was fitted.

4.2 The Annotation on Empathic State

4.2.1 Annotation

In our study, we used the speech analysis tool Praat, and we conducted speech analysis, labeling, annotation and the calculation of basic sound parameter. A consecutive continued speech (same person) was defined as a speech segment, and a consecutive continued topic defined as a topic section. 1290 speech segments and 264 topic sections were obtained from all data.

We set three kinds of label items that related to empathic statuses: the topic, the attitude and the emotional state.

- The Topic: as for this item, it is labeled to identify whether a new topic is raised or not. Two cases are included. One is that a new topic is raised, and the other is that the topic is converted to a related one.
- The Attitude: In this study, the attitude is labeled in three basic categories, namely "positive", "negative" and "neutral". Present studies [3] often focus on the reaction after the "question and answer", and analyze the questioner's attitudes towards the responder. However, no such limitation is found in this study, as all speech segments are analyzed.
- The Emotional State: the classification of emotional state adopts basic categories for emotion which are proposed by Ekman [2], and a state of "calm" is added.3.2.1 Summary of annotation results

In addition, we considered label of "equivocal" "unknown": "equivocal" was used to labeling the case of no appropriate category for select (both the emotional states and attitudes). "Unknown" was used to labeling the case of the difficult to classify on the state of sound. We use the software Praat to tag the speech segments.

4.2.2 Summary of Annotation Results

The total 1290 speech segments were labeled by three persons. The label classification was determined by the majority vote, and failing to get the majority vote was classified as "mismatch".

- The Topic: topic raised = "absence" is 79%, topic raised = "presence" is 21%. Every five utterances have one topic raised on average. However, the analysis of variation has not been performed yet. "Mismatch" is few, so "the topic" is well labeled.
- The emotional state: "calm" (62%), "joy" (19%), "surprise" (15%), appear of other emotional states hardly happens. This is due to characteristics of our dialogue data. "Mismatch" is relatively little, which is just 2%. It means the labeling of emotional state ("joy" and "surprise") is fairly stable.
- The attitude: "neutral" is 69%, "positive" is 20%, and "negative" is relatively little, just 4%. Because, topics of our dialogues rarely cause disagreement. On the other hand, "unknown" is 4%, it suggests that three classification of attitudes maybe insufficient. "Mismatch" is relatively little, just 2%. It means labeling of attitude is fairly stable too.

5. Speech Analysis of the Speech Segment

In order to make clear of the speech feature in each empathic state, please refer to previous studies ([Banse 1996], [Cornelius 1996], which have performed the basic speech analysis. In this analysis, seven parameters (three types) below were used: The sound pressure and fundamental frequency were analyzed by audio signal analysis function of Praat.

- Pitch (fundamental frequency F0): maximum, minimum, average value
- Loudness (sound pressure): maximum, minimum, average
- Speech speed (Mora per second): calculated from the period of speech segment and the romanization transcriptions

In previous studies, it is found that the speech voice is affected by characteristics of individual and gender [Nishio 2005 2004]. So the value of each speech parameter is set in the basic state [attitude = neutral, emotional state = calm] as reference value, and speech variations are calculated on the basic state (standard value). Here, in order to minimize the influence of personal characteristics, the ration of fundamental frequency and speech speed are calculated, as well as the sound pressure different from the basic state of each person

6. Estimation of Empathic State with Machine Learning

6.1 Identification Process

The estimation of "attitude" and "emotion" of speech segments is based on the definition of "empathic state" in this study and the knowledge obtained from previous studies. Therefore, we propose the identification process of multi-tiered classifier as shown in Fig 1 below. [Flow of dialogue] in Figure 1 indicates the local structure of the dialogue in this study.

Speech segment Y is a respond to speech segment X, unless it raises a new topic. In Figure 1, the direction of

the arrow indicates the reaction to which speech segment. Various classifiers are arranged in the identification process. There are three classifiers: classifier of a new topic, classifier of "attitude" and classifier of "emotion". Classifier of "attitude" and "emotion" is further subdivided, it is caused by some speech segments which do not react to any speech segment and some times the identification result of preceding speech segment can be used.

At the beginning of the identification process, the classifier new topic submission identifies the target utterance Y whether or not the speech of a new topic submission first. Based on this we can recognize the topic section. It is based on the research results that a new topic is significantly different phonetically from other "empathic states" [Gaoyahan, 2010b]. Then, based on this identification result, classifier attitude and emotion estimates the respective categories. Here, given that probably the identification of the preceding speech X has been completed. In this case, use of the identification result of X for the identification of Y can be considered. Further, when identifying Y, the attitude can be identified first, then use such result to identify the emotion. The reverse is also contemplated. This known information is referred to as additional information in the identification process.

6.2 Learning of Classifiers by Machine Learning

The category data obtained from annotation and the data of speech feature were used as learning data learned each classifier described below by machine learning (SVM: Support Vector Machine).

The 2128 speech segments (men: 1218, women: 910) were used for the study, as "indistinguishable" and deficient "unknown" data were screened out from the. original 2241 speech segments (men: 1294, women: 947). As for the emotion category, it is limited to "calm", "joy" and "surprise"

6.2.1 Classifier of New Topic Submission

It will be identified if the speech segment Y is a new topic submission or not. Input information is only speech feature. Using this classifier can identify the topic section.

6.2.2 Classifier of Emotion

It is divided into three cases according to the information of the target reaction.

- If speech segment Y does not react to any preceding speech segment: This is a speech segment of the new topic submission. Just identify emotion category by using speech features of the speech segment Y.
- If the speech segment Y is not a respond to a new topic submission X: other than the speech feature information of X, there are categories of "emotion" that can be used as additional information for identification. Moreover, if the emotion category of Y is identified before, this can be used as additional information as well.
- If the speech segment Y reacting to speech segment X is not a new topic submission: other than the speech feature information of X, there are categories of "emotion" and "attitude" that can be used as additional information for identification. Moreover,

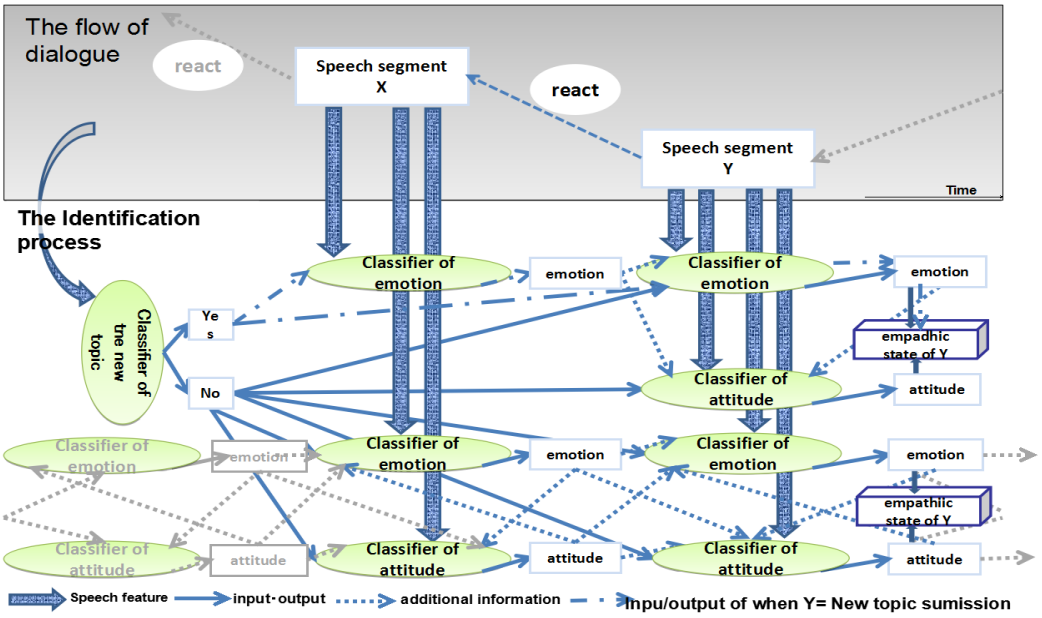

Figure 1: Identification process of the empathic state

if the attitude category of Y is identified before, this can be used as additional information as well.

6.2.3 Classifier of Attitude

This is the same with classifier of emotion, divided into three cases according to the information of the target reaction.

- If the speech segment Y is not a respond to any preceding speech segment: there is no need to identify the attitude in this case in this study, according to our definition of "attitude".
- If the speech segment Y is a respond to a new topic submission X: other than the speech feature information of X, there are categories of "emotion" that can be used as additional information for identification. Moreover, if the emotion category of Y is identified before, this can be used as additional information as well.
- If the speech segment Y reacting to speech segment X is not a new topic submission: other than the speech feature information of X, there are categories of "emotion" and "attitude" can be used as additional information for identification. Moreover, if the emotion category of Y is identified before, this can be used as additional information as well.

6.3 Results of Identification and Discussion

6.3.1 Identification Result

In this study, the amount of annotation data is limited. So we used cross-validation of the three portions, use error rare to evaluate the classification accuracy. Analyze the frequency distribution of categories in learning data. Take the most frequent category as baseline to compare with the identification result. In the following, the result of learning is shown. This is a result based on data from the research involving male and female students.

6.3.2 Identification Result of New Topic Submission Speech

The error rate of all data is 0.0395 (deviation from the baseline: 0.001), 0.0462 (0.001) in men, 0.0292 (-0.003) in women. It is suggested that, in order to improve the detection accuracy of the new topic submission, more information is required (The time interval between speech segments, cues in speech content, and information from other modalities such as gesture or line of sight etc.). In addition, the result of women is slightly better than that of men, but from the fact that the identification result is not good even if two sets of data of men and women are matched. It is suggested that speech expression of new topic submission is different between men and women.

6.3.3 Identification Result of Emotion and Attitude

In the identification process, the input information is different which depends on whether the speech segment of X or Y is a new topic submission. Therefore, we divided three cases as below: (1) No matter whether X is a new topic submission or not, Y is a new topic submission. (2) X = new topic submission, and Y = not new topic submission, (3) X = not new topic submission, and Y = not new topic submission. The above three cases are used to analyze the relationship between result of identification and input information

(1) No matter whether X is a new topic submission or not, Y is a new topic submission.

In this case, there is no need to identify attitude (not defined in this case) of Y, but it is required to identify the emotion category of Y. Input information is speech features of Y only. The results of error rate is 0.1692 (-0.081) in total, 0.1351 (-0.065) in men, and 0.3333 (-0.011) in women. Both lower than the baseline, the effectiveness of the machine learning was confirmed. In comparison with men, the identification results of women are worse, it is considered that categories to identify emotions are more than men, and the number of learning data was small.

(2)X = new topic submission, and Y = not new topic submission,

In this case, other than the speech feature of Y, the emotion category of preceding speech segment X that classified in advance is available. Since X is a new topic submitted speech, the identification result of the attitude category is not available. However, one of the emotion or attitude of the identification target utterance Y is identified in advance; this identification information can be used as additional information to identify another category of this. The identification result shown in Table 2 is the identification error rate. Values in parentheses indicate the difference from baseline. (Result of identifying only use speech features.) That means a negative value indicates that the result of the classifier constructed by our learning data is better than the baseline classifier. Bold values in the table represent the lowest identification error rate in each identification condition.

Table 2: Identification result of empathic state in case (2)

Target	Input information	Man	Woman	Total
Attitude of Y	Speech feature only (▲)	0.1256	0.2242	**0.1667**
	▲+emotion of X	0.1324 (0.0068)	0.2639 (0.0397)	0.172 (0.0053)
	▲+emotion of Y	0.125 (-0.0006)	**0.1939 (-0.0303)**	**0.1667 (0)**
	▲+emotion of X + emotion of Y	**0.1186 (-0.007)**	0.2593 (0.351)	0.172 (0.0053)
Emotion of Y	Speech feature only (▲)	0.2790	**0.1606**	**0.2353**
	▲+emotion of X	0.2253 (-0.0537)	0.2315 (0.0709)	0.2581 (0.0228)
	▲+attitude of Y	0.2524 (-0.0266)	0.2212 (0.0606)	**0.2553 (0)**
	▲+emotion of X + attitude of Y	**0.1812 (-0.0978)**	0.2269 (0.0663)	0.2366 (0.0013)

From table 2 the following information can be concluded.

Identification of attitude:

- Although using emotion category of X as additional information, the error rate is not lower than the baseline.
- When assuming that emotion category of Y is identified in advance and using this as additional information, the error rate in both men and women were decreased from baseline. But, it does not decrease from the baseline in matched data (men and women)
- In addition to the above conditions, when introducing the emotion category of X in as additional information, identification error rate decreases from baseline in men, but reduction was not observed in women. It does not decrease from baseline when used in matched data (The data which is not divided by gender).

Identification of emotion:

- When attitude category of Y is used as additional information, identification error rate is reduced from baseline only in men.
- When using emotion category of X and attitude category of Y as additional information, identification error rate is reduced from baseline only in men.
- For women, the identification error rate of the results using only the speech feature is the lowest, reduction of the identification error rate was not observed when using additional information known category.

From the above results, the following conclusions can be drawn:

- In both men and women, there is a tendency that identification error rate is lower in the identification of attitudes than emotion.
- Use of the additional information is effective, but the effect is limited.
- For men: The identification error rate of attitude is lower than that of emotion.

In the identification of emotion category, when attitude category is identified earlier than emotion category, and this result is combined with emotion category of X to be identified, the whole identification error rate will be improved.

- For women: in contrast to men, there is a tendency that the identification error rate of emotion is lower than that of attitude.

When identifying the emotion category of Y earlier than attitude and using this as additional information to identify attitude, the identification error rate will be improved.

(3) X = not new topic submission, and Y = not new topic submission.

In this case, both attitude and emotion of X was defined, so definition about these can be used as additional information for identification. It is the same with case (2), if any one of emotion category or attitude category was identified in advance, it could be used as additional information for identification of another one. The identification result is shown in Table 3.

Table 3: Identification result of empathic state in case (3)

Target	Input information	Man	Woman	Total
Attitude of Y	Speech feature only (▲)	0.2525	0.2356	0.2464
	▲+emotion of X	0.2533 (0.0008)	0.2404 (0.0048)	0.2484 (0.002)
	▲+attitude of X	0.2525 (0)	0.2369 (0.0013)	0.2458 (-0.0006)
	▲+emotion of X + attitude of X	0.2544 (0.0019)	0.2404 (0.0048)	0.2491 (0.0027)
	▲+emotion of Y	0.2556 (0.0031)	0.2295 (-0.0061)	0.2444 (-0.002)
	▲+emotion of X + emotion of Y	0.2564 (0.0039)	0.2391 (0.0035)	0.2477 (0.0013)
	▲+attitude of X + emotion of Y	0.2556 (0.0031)	0.2213 (-0.0143)	0.245 (-0.0014)
	▲+emotion of X+ attitude of X+emotion of Y	0.2553 (0.0028)	0.239 (0.0034)	0.2484 (0.002)
Emotion of Y	Speech feature only (▲)	0.2588	0.314	0.2882
	▲+emotion of X	0.2602 (0.0014)	0.3071 (-0.0069)	0.2896 (0.0014)
	▲+attitude of X	0.2599 (0.0011)	0.3302 (0.0162)	0.2922 (0.004)
	▲+emotion of X + attitude of X	0.2646 (0.0058)	0.3145 (0.0005)	0.2883 (0.0001)
	▲+attitude of Y	0.2649 (0.0061)	0.3101 (-0.0039)	0.2878 (-0.0004)
	▲+emotion of X + attitude of Y	0.271 (0.0122)	0.3008 (-0.0132)	0.2882 (0)
	▲+attitude of X + attitude of Y	0.2618 (0.003)	0.3251 (0.0111)	0.2937 (0.0055)
	▲+emotion of X+ attitude of X+ attitude of Y	0.2721 (0.0133)	0.3038 (-0.0102)	0.2908 (0.0026)

The following conclusions can be drawn from Table 3.

Identification of attitude:

- In men, the identification error rate is the lowest when identified in only speech feature.
- For women, when using the attitude category of X and emotion category of Y as additional information, the identification error rate is lowest.
- When using emotion category of X as additional information in both men and women, the identification error rate is not decreased from the baseline. When using emotion category only as additional information in the matched data of men and women, the identification error rate is lowest.
- When using all available additional information, identification in the data of men, women or their matched data, error rate does not decrease from the baseline.

Identification of emotion:

- For men, the identification error rate is the lowest when identified in only speech feature.
- For women, when using the emotion category of X and attitude category of Y as additional information, the identification error rate is lowest. It has a symmetric relationship with identification of attitude.
- When using attitude category of X in both men and women as additional information, the identification error rate is not decreased from the baseline. When matched the data of men and women, the identification error rate is lowest when using attitude category of Y as additional information only. This also has a symmetric relationship with the identification of attitude.
- When using all available additional information, identification in the data of men, women or their matched data, error rate does not decrease from the baseline. This result is the same with the identification of attitude.

From the above results, some trends can be identified, as follows.

- there is a tendency that the identification error rate is lower in the identification of attitudes than emotion in both men and women
- Use of the additional information is more limited than identifying the attitude. But, there is a tendency that using additional information is slightly better for women.

If the attitude category of Y is identified before emotion category of this, and this result is combined with emotion category of X as additional information to identify emotion category, the identification error rate will be improved.

7. Discussion

7.1 The Speech Feature of Emotion and Attitude in the Identification

Based on these result, a trend can be identified that the identification error rate of the attitude category is lower than that of the emotion category. It seems to contradict the knowledge we have gained from researches so far, which is "speech characteristic appear on emotion many than attitude" [Gaoyahan, 2010b].

Findings from the above study:

(1) The types of speech parameter are more significantly different between emotions at the same attitude compared to those between attitudes at the same emotion.

(2) It is based on the assumption that the width of the T-value at the speech parameters of the same type is larger between emotions than between attitudes. (For example, the T-values of the sound pressure maximum value between attitude: 4.95 to 6.35, same parameter's T-value between emotion: 6.58 to 15.00). This is due to the use of all the seven speech parameters in this identification process, resulting in a contradictory to the above findings, however, further studies are required.

7.2 Difference of the Speech Expression in Man and Woman

From our previous research, it is concluded that "speech expression of empathic state is different between man and women "[Gaoyahan, 2010b]. The results from this study identify a tendency that the result of identification from data divided in men and women is better than in matched data of men and women. It is supported by the result from our previous research. When analyze it in more detail in the identification of emotion and attitude, a symmetrical result has also been observed between men and women. Further analysis need to be done on this.

7.3 Effectiveness of the Additional Information

Our results suggest some extent of effectiveness of the additional information, but it is a result within a very limited range. This can be improved by using more diversified speech parameters. On the other hand, it is conceivable to make clear of the dependencies in target speech with preceding speech at empathic states, and use this in the identification process. For example, if it is clear that the speech segment is reacted to a certain piece of preceding speech segment, the reduction of the identification error rate of about 1% is observed. It indicates that there are some dependencies between empathic states of analyze target speech segment and which preceding speech segment it responds to.

8. Conclusion

This paper examines a method to estimate empathic states from speech utterances. The idea is to use the annotated data as training data for a machine learning algorithm (SVM) to construct relevant classifiers. These classifiers are arranged in such a way to reflect our definition of empathic states. In this paper, some interesting findings have been obtained based on analyzing the additional information which is assumed to be known already, but significant improvement from baseline has not been found. In the future, it is necessary to examine the dependency of empathic states between speech segments in the response relationship from aspects of dialogue while considering the diversification of speech parameters. In addition, information of other modes, such as images, will be adopted as well to improve the detection accuracy of the new topic submission.

References

[1]R. Banse, K. R.Scherer, Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology，70(3),* 1996, 614-636 .

[2] R.R Cornelius, The Science of Emotion: Research and Tradition in the Psychology of Emotions. *Pearson Education* ,1996.
[3]P. Ekman, An argument for basic emotion. *Cognition and Emotion*, *6(3/4)*, 1992, 169-200.
[4]A. Mehrabian, N. Epstein, A measure of emotional empathy. *Journal of Personality*, *40(3)*, 1972, 525-543.
[5] I.R, .Murray, J.L. Arnott, Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *Journal of the Acoustic Society of America*, 93(2), 1993, 1097-1108.
[6]Gaoyahan, Y. Ohno, Extraction of speech parameters relating to the characteristics of emotional expression: Focus on the attentiveness-related parameters such as the nursing experience, The Japanese Journal of Applied IT Healthcare, 5(2), 2010, 114-123.
[7]Gaoyahan, Y.Ohno, 2D-6 Fundamental analysis of empathic statuses in dialogues, *Information Processing Society of Japan (IPSJ)*, 2D-6, (2010a).
[8]Gaoyahan, Y. Ohno, Speech Analysis of Empathic States in Spontaneous Utterances. *Forum on Information Technology2010*, RJ-002 (2010b).
[9]T. Maeda, K. Takashima, A Study of Nonverbal Cues and Atmosphere in Three-Person Conversation. *IEICE technical report*, 2010.