IEICE TRANS. INF. & SYST., VOL.E95–D, NO.11 NOVEMBER 2012

PAPER

A Method for Predicting Stressed Words in Teaching Materials for English Jazz Chants

Ryo NAGATA[†a)], ***Member***, **Kotaro FUNAKOSHI**[††], **Tatsuya KITAMURA**[†], ***Nonmembers***, ***and*** **Mikio NAKANO**[††], ***Member***

SUMMARY To acquire a second language, one must develop an ear and tongue for the correct stress and intonation patterns of that language. In English language teaching, there is an effective method called *Jazz Chants* for working on the sound system. In this paper, we propose a method for predicting stressed words, which play a crucial role in Jazz Chants. The proposed method is specially designed for stress prediction in Jazz chants. It exploits several sources of information including words, POSs, sentence types, and the constraint on the number of stressed words in a chant text. Experiments show that the proposed method achieves an F-measure of 0.939 and outperforms the other methods implemented for comparison. The proposed method is expected to be useful in supporting non-native teachers of English when they teach chants to students and create chant texts with stress marks from arbitrary texts.

key words: *language learning, stress prediction, teaching material generation, Jazz Chants, stress-timed rhythm*

1. Introduction

To acquire a spoken language, one must develop an ear and tongue for the correct stress and intonation patterns of the spoken language. This is normally difficult for those who are acquiring a second language whose sound system is not similar to that of their first language. An example pair would be English and Japanese in which the sound systems are quite different.

In English language teaching, there is an effective method called *Jazz Chants*[*] for working on the sound system. "A chant is a rhythmic expression of natural language which links the rhythms of spoken American English to the rhythms of traditional American jazz — the rhythm, stress and intonation pattern of what children would hear from an educated native speaker in natural conversation [1]". In chants, each stressed word is pronounced (i) with an extra emphasis[**] (often with physical activities such as clapping or jumping) and (ii) with an equal time interval (i.e., isochronism). To support this, stressed words are sometimes (but not always) marked with the asterisk * or underlined in teaching materials for chants (Hereafter, teaching materials for chants will be referred to as *chant texts*). An example of a chant text is as follows [1]:

* * * *
Frank, Hank, walk to the bank.

* * * *
Jill, Phil, run up the hill.

Teachers and children read chant texts out loud, putting stress on the marked words. Note that the time interval between *Frank* and *Hank* and that between *walk to the* and *bank* are equal although the latter has more words (tree words) than the former does, which means that the latter is pronounced more quickly than the former is.

Since chants require only sound and physical activities to teach, they are especially suitable for children who are not yet familiar with written language. In addition, Graham [1] shows that the use of chants has the following three advantages in language learning and teaching:

1. Acquiring stress and intonation patterns
2. Memorizing everyday phrases
3. Learning grammar and vocabulary

At the same time, the use of chants has a drawback for non-native speakers of English. It is crucial to recognize stressed words in chants. However, chant texts often do not mark stressed words because chants were originally designed for teachers who are native-speakers of English and who naturally recognize where to place the stresses. By contrast, non-native speakers of English, even teachers of English, have difficulties in recognizing stressed words in some cases. For instance, those who were not originally teachers of English but of other subjects are now in charge of English language teaching in primary schools in Japan. To reduce this difficulty, it is preferable that teaching materials for chants should explicitly mark stressed words for non-native teachers of English as well as for learners of English.

In order to predict stresses in chants, one could apply conventional pitch-accent prediction methods such as [2], [3]. However, the conventional pitch-accent prediction methods normally require acoustic information which is not available stress prediction for chants. More importantly, although stresses in chants share similar properties with pitch accents, they seem not to be identical. Stresses in a chant text have special properties as will be described in Sect. 2. It

Manuscript received May 17, 2012.

[†]The authors are with Konan University, Kobe-shi, 658–8501 Japan.

[††]The authors are with Honda Research Institute Japan Co., Ltd., Wako-shi, 351–0188 Japan.

a) E-mail: rnagata@konan-u.ac.jp

DOI: 10.1587/transinf.E95.D.2658

[*]Jazz Chants is a registered trademark of Oxford University Press. In this paper, Jazz Chants will be simply referred to as *chants*.

[**]In chants, each stressed word is somewhat exaggeratedly pronounced to acquire the rhythm, stress and intonation patterns.

is likely that one will have to modify the conventional pitch-accent prediction methods to achieve a good performance in stress prediction in chants. Nagata et al. [4] investigated how well a simple Hidden Markov Model (HMM) based method works on stress prediction in chants. They showed that the problem can be solved as a sequence labeling problem using HMMs where the input is a sequence of words or part-of-speech (POS) tags obtained from the chant text in question. At the same time, it was argued that, in stress prediction for chants, it is crucial to consider the properties of chants such as a constraint on the number of stressed words in a chant text, which will discussed in Sect. 2.

Accordingly, we propose a stress prediction method specially designed for chants. This method exploits several sources of information including words, POSs, sentence types, and the constraint on the number of stressed words, which are relevant in stress prediction for chants. The proposed method is expected to be useful in supporting non-native teachers of English when they teach chants to students; it can provide them with the information about which word gets stressed in a given chant text. It should also be useful for them to create their own teaching materials, which teachers often do. Note that it is often the case that native speakers of English are not readily available in certain countries including Japan. In addition to supporting teachers, it can be applied to software programs for learning the English sound system. For example, it can be used to instruct learners which word gets stressed in a given chant text. Ultimately, it can be applied to a chanting robot that interactively teaches the English sound system based on chants as Nagata et al. [4] originally proposed. It is crucial for such robots to recognize stressed words in the utterances.

The rest of this paper is structured as follows. Section 2 explores chants in more detail, which is necessary to discuss the proposed method. Section 3 describes the proposed method. Section 4 describes experiments conducted to evaluate the proposed method. Section 5 discusses the experimental results.

2. Looking into Chants

There are some basic tendencies in which words get stressed in chants. Content words such as nouns and verbs tend to get stressed more often than function words such as determiners and prepositions. This implies that information on POSs is crucial for stress prediction. Also, information on words plays an important role since some of the words that fall into the same POS category get stressed and others do not. For example, while the words *you* and *it* fall into the same category *pronoun*, the former tends to get stressed more often than the latter. Therefore, information on both words and POSs needs to be considered in stress prediction.

One factor which is not as obvious as words and POSs is sentence types. In questions, interrogatives such as *where* and sometimes auxiliaries such as *does* get stressed as in *Where is my hat?*. Correlated with this is the relation between sentence types. The determination of stressed words in a sentence is sometimes influenced by the type of its previous sentence. For example, if the previous sentence is a *where*-question as in the above example, one of the prepositions in the next sentence is likely to get stressed (e.g., *It's on the table.*).

Another important factor is the constraint on the number of stresses in a chant text; it is constrained to be a multiple of eight. This may seem to be somewhat odd, but is explained as follows. Chants are normally performed with music that progresses regularly in 4/4 time (recall that chants are formally *Jazz Chants*) where each beat corresponds to each stressed word. Music is often based on two bars (i.e., motive), which consists of eight beats in 4/4 time, or their multiples (e.g., 16 beats in four bars, 24 beats in six bars, ...). Consequently, the number of stresses in a chant text is constrained to be a multiple of eight. It should be emphasized that null stressed words are sometimes inserted in a chant text to satisfy the constraint (e.g., "*Black, yellow, brown. NULL. Jack fell down. NULL*" [1] where *NULL* denotes a null stressed word). Null stressed words are not actually pronounced but can be expressed with physical activities such as a clap.

3. Proposed Method

The stress prediction task can be solved as a sequence labeling problem. The sequence of observed values is the sequence of words in a given chant text. The labels are binary and denote whether the word gets stressed or not. Take for example a sentence in the textbook for chants [1]:

* * * *
Frank, Hank, walk to the bank.

This can be alternatively expressed with a sequence of labels *S* and *N*:

Frank/S, Hank/S, walk/S to/N the/N bank/S.

where S and N denote stress and not-stress, respectively (hereafter, S and N will be used to denote stress and not-stress).

To solve the sequence labeling problem, we use conditional random fields (CRFs) [5], which have been shown to be effective in sequence labeling. One of the reasons why we use CRFs is that it can handle several sources of information. As discussed in Sect. 2, the determination of stressed words in chants relies on several factors including information on words, POSs, and sentence types. Also, as we will see bellow, CRFs have several favorable properties in stress prediction.

To define the stress prediction method based on CRFs, we will use the following symbols. We will denote the sequence of observed values and the sequence of corresponding labels by $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively. In other words, $\boldsymbol{x}$ and $\boldsymbol{y}$ refer to the sequence of input words and the corresponding sequence of S or N, respectively. We will also denote a feature function by $\phi(\boldsymbol{x}, \boldsymbol{y})$. In our methods, the feature functions are binary-valued. For instance, one of the feature

function would be a function that returns 1 if the word question in $\boldsymbol{x}$ is "walk" and the corresponding label is S, and 0 otherwise (we will shortly describe the actually used feature functions below). In our methods, we will limit ourselves to first-order Markov model features to encode inter-label dependencies. Also, we will denote all feature functions by the feature function vector $\boldsymbol{\phi}(\boldsymbol{x}, \boldsymbol{y})$.

With these symbols, the probability of the label sequence $\boldsymbol{y}$ given the word sequence $\boldsymbol{x}$ is calculated by

$$p(\boldsymbol{y}|\boldsymbol{x}) = \frac{\exp\{\boldsymbol{w}^t\boldsymbol{\phi}(\boldsymbol{x}, \boldsymbol{y})\}}{\sum_y \exp\{\boldsymbol{w}^t\boldsymbol{\phi}(\boldsymbol{x}, \boldsymbol{y})\}} \tag{1}$$

in CRFs where $\boldsymbol{w}$ denotes weights for feature functions. The weights are estimated by using training data.

We use four types of features in the feature functions: (i) words, (ii) lemmas of words, (iii) POSs, and (iv) sentence types. For (i) to (iii), we set the window size to five: current word, two previous words, and two following words. In addition, we include bi-grams and tri-grams extracted from the window: bi-grams consisting of the previous word and the current word, and the current word and the following word; tri-grams consisting of the previous word, the current word, and the following word. For (iv), we consider the combinations of the type of the sentence in which the current word appears and that of the previous sentence; the sentence types used are declarative, *yes/no*-question, *what*-question, *where*-question, *when*-question, *who*-question, *why*-question, and *how*-question. These are the features we use in the proposed method.

With CRFs and these features, we can make basic predictions. First, we break down the input chant text into feature vectors. Then, we put the feature vectors into CRFs to obtain predictions and the corresponding probabilities. Namely, we simply select $\boldsymbol{y}^*$ given by

$$\boldsymbol{y}^* = \arg\max_{\boldsymbol{y}} p(\boldsymbol{y}|\boldsymbol{x}) \tag{2}$$

Alternatively, we can obtain the N-best label sequences according to the probabilities given by Eq. (1).

To satisfy the constraint on the number of stresses in a chant text, we can exploit the conditional probabilities given by CRFs. We search the N-best prediction results for the label sequences that satisfy the constraint. In other words, we count the number of stressed words in each predicted sequence and consider those whose number is a multiple of eight. Among them, we can simply choose the one that maximizes the conditional probability as the prediction result, which is another advantage of using CRFs. Formally, we select $\boldsymbol{y}^*$ given by

$$\boldsymbol{y}^* = \arg\max_{\boldsymbol{y}\in\{\boldsymbol{y}|n(\boldsymbol{y})=8i, i\in\mathbb{N}\}} p(\boldsymbol{y}|\boldsymbol{x}) \tag{3}$$

where $n(\boldsymbol{y})$ denotes the number of S in the sequence $\boldsymbol{y}$.

In addition to the constraint, we consider the distribution of the length between stress-intervals. Here, we define a stress-interval as an interval between a stressed word and the word before the next stressed word[†]. For example, there are three stress-intervals *Frank*, *Hank*, and *walk to the* in *Frank/S, Hank/S, walk/S to/N the/N bank/S*. Theoretically, one can put as many words as one wants in a stress-interval in English. Practically, however, too many words in a stress-interval (or too long stress-interval) make it difficult to pronounce the stress-interval properly. Accordingly, the length of stress-interval is expected to be distributed among certain lengths. In other words, there might be a prediction error in a too long stress-interval predicted by CRFs.

To consider the distribution of the length of stress-intervals, we have to solve two technical problems: (1) how to measure the length of stress-intervals and (2) how to combine the distribution with CRFs. In this paper, we measure the length of a stress-interval by the number of not-stressed words in it[††]. For example, the length of the stress-interval *walk to the* is two. To solve the second problem, we assume that the length follows the Poisson distribution, which gives the probability of the number of events occurring in a fixed time. Namely, the number of events in a fixed time corresponds to the number of not-stressed words appearing in a stress-interval, which in turn corresponds to the length of a stress-interval. Under this assumption, we can calculate the probability of the length of a stress-interval once we estimate the mean of the length. Then, we can combine it with CRFs by simply multiplying both probabilities (this is another reason why we use CRFs).

We calculate the probability of the length of stress-intervals as follows. We first estimate the mean of the length by using the training data. To formalize the calculation, we will denote the number of stress-intervals in the training data as M. Also we will denote the length of the m-th stress-interval as l_m $(= 0, 1, 2, \ldots, M)$ in the training data. Then, we estimate the mean by:

$$\lambda = \frac{1}{M}\sum_{m=1}^{M} l_m. \tag{4}$$

Using Eq. (4), we can calculate the probability of the length l following the Poisson distribution with parameter λ by:

$$f(l) = \frac{\lambda^l}{l!}\mathrm{e}^{-\lambda}. \tag{5}$$

Since we have M stress-intervals in a chant text, we take the geometric mean of the probabilities, which is given by:

$$\bar{f}(\boldsymbol{y}) = \sqrt[|S_y|]{\prod_{s\in S_y} f(l_s)}. \tag{6}$$

This value can be interpreted as a score that evaluates how

[†]If the first word in a chant text is not stressed, then the stress-interval is between the first word and the word before the first stressed word. Similarly, if the last word is not stressed, then the stress-interval is between the last stressed word and the last word.

[††]We also used the number of syllables in not-stressed words instead of the number of not-stressed words. However, it did not make any difference in the prediction performance. Therefore, we selected the number of not-stressed words as the length of stress-intervals, which is much easier to count.

good a prediction result is, solely relying on the length of stress-intervals.

To make the final prediction, we combine the geometric mean with the prediction results obtained by the CRFs. For the N-best results obtained by the CRFs that satisfy the constraint on the number of stressed words, we calculate

$$\boldsymbol{y}^* = \arg \max_{\boldsymbol{y} \in \{\boldsymbol{y} | n(\boldsymbol{y}) = 8 i, i \in \mathbb{N}\}} p(\boldsymbol{y}|\boldsymbol{x}) \bar{f}(\boldsymbol{y}) \quad (7)$$

where p, $\bar{f}$, and $n(\boldsymbol{y})$ denote the conditional probability given by the CRFs (Eq. (1)), the score based on the length of stress-intervals (Eq. (6)), and the number of S in the sequence $\boldsymbol{y}$, respectively. We choose the one that maximizes Eq. (7) as the final prediction. In other words, we make a prediction considering features around the word in question, the constraint on the number of stressed words, and the length of stress-intervals as required in chant texts. If no predicted sequences satisfy the condition, we simply choose the sequence that maximizes the probability obtained by the CRFs.

4. Experiments

For evaluation, we used 71 chant texts in the textbook [1]†, which are manually annotated with stresses. In the experiments, we assumed that null stressed words were given and we excluded them from the evaluation. We used a POS tagger†† to annotate the chant texts with POS tags. In all, the 71 chant texts consisted of 2,396 tokens and 1,531 stressed words.

To measure the performance, we used recall, precision, F-measure, and accuracy. Recall and precision were defined by

$$R = \frac{\text{Number of stressed words correctly predicted}}{\text{Number of stressed words}} \quad (8)$$

and

$$P = \frac{\text{Number of stressed words correctly predicted}}{\text{Number of words predicted to be stressed}}, \quad (9)$$

respectively. F-measure was defined by

$$F = \frac{2RP}{R+P}. \quad (10)$$

Accuracy was defined by

$$A = \frac{\text{Number of chant texts without prediction error}}{\text{Number of chant texts}}. \quad (11)$$

All measures were calculated by leave-one-out cross-validation [6] (one text was left out each time).

In the experiments, we implemented seven methods including previous methods for comparison. The first is a baseline where all tokens are predicted to be S (Baseline). The second is the previous method [4] based on the POS tri-gram HMMs (HMM)†††. The third and fourth are based on CRFs, but use only word features and POS features, the same used in the proposed method, respectively (CRF word only and CRF POS only). The fifth is a CRF-based method exploiting all the proposed features but without the constraint on the number of stressed words and the length distribution (CRF base). The sixth is the CRF-based method with the constraint on the number of stressed words (CRF constraint). The value of N (N-best prediction results) was set to 10. The seventh is the CRF-based method with the constraint on the number of stressed words and the length distribution (CRF all). Again, the value of N was set to 10.

Table 1 Experimental results.

Method	R	P	F	A
Baseline	**1.00**	0.639	0.780	0.281
HMM POS	0.914	0.853	0.883	0.423
CRF word only	0.915	0.893	0.904	0.451
CRF POS only	0.933	0.903	0.918	0.465
CRF base	0.946	0.926	0.936	0.507
CRF constraint	0.950	**0.927**	**0.939**	**0.592**
CRF all	0.949	0.926	0.937	0.535

R: Recall, P: Precision, F: F-measure, A: Accuracy

Table 1 shows the experimental results. It reveals that the proposed methods outperform the baseline and the previous method. It also reveals that the CRF-based methods tend to improve as the available information increases. The next section compares the proposed methods in more detail.

5. Discussion

5.1 Comparison of Methods

As Table 1 reveals, all CRF-based methods outperform the HMM-based method. Even "CRF word only" or "CRF POS only" perform better than the HMM-based method does. This is because that CRF-based methods exploit information before and after the word in question including bi-gram and tri-gram features unlike the HMM-based method. The CRF-base method further improves when it combines POS features with word features as we expected. Basically, POSs are informative for determining which word to stress as Table 1 shows; "CRF POS only" performs better than "CRF word only" does. However, information on words are required in some cases. For instance, the word *I* tend to get stressed and the word *it* do not although both fall into the same POS category *pronoun*. In other words, it is crucial to exploit both the sources of information in stress prediction.

The CRF-based method performs very well when it exploits all the proposed features as "CRF base" shows. It further improves when it considers the constraint on the number of stressed words. As already explained, chant texts tend to satisfy the constraint on the number of stressed words and "CRF constraint" (and "CRF all") make prediction sat-

†We corrected some stress marks, which seemed to be typographical errors in three chant texts out of the 71 by consulting a professional chants trainer and the accompanying CD.

††http://nlp.ii.konan-u.ac.jp/tools/hmmtagger/index.html

†††We chose the POS tri-gram HMMs because they perform better than the word tri-gram HMMs according to Nagata et al. [4].

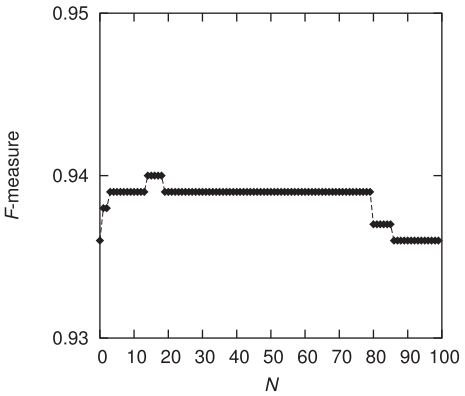

Fig. 1 Relation between parameter N and performance (F-measure).

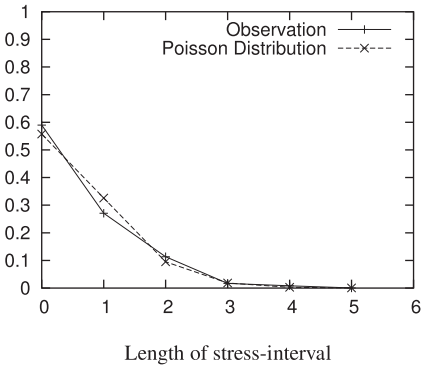

Fig. 2 Distribution of length of stress-intervals.

isfying the constraint. As a result, it achieves the best F-measure and accuracy. Beside, around 60% of given texts are expected to require no modification to the prediction results, which is an advantage for teachers who use the prediction results. These results suggest that it is crucial to consider the constraint on the number of stressed words in stress prediction for chants as we discussed in Sect. 2.

A drawback to "CRF constraint" is that it has the extra parameter N for the N-best label sequences to be searched. One needs to find its optimal value to achieve good performance, which often requires an additional data set (development data). To investigate the relationships between the performance and N, we conducted additional experiments with $N = 1, 2, \cdots, 100$. It turned out that although "CRF constraint" did improve as the value of N increased, the improvement soon converged after N exceeded a certain value as Fig. 1 shows. Considering these results, it follows that (i) one should estimate the optimal value of N using development data if sufficient training data are available; (ii) if it is not the case, one should set N to a small value (e.g., $N = 10$), which achieves better performance than CRF-based methods without the constraint, reducing the computational cost.

Contrary to our expectation, the distribution of the length between stress-intervals did not contribute to further improving the performance. On the one hand, a closer look at the experimental results revealed that the distribution closely followed the Poisson distribution as we had expected; Figure 2 shows the distribution of the length of stress-intervals observed in the experimental data and the Poisson distribution whose mean was estimated from the experimental data, respectively. On the other hand, the CRF-based methods can achieve highly good performance even without the length distribution. Thus, there is only very little room for improvement. In other words, it was not often the case that the prediction results contained too short or too long stress-intervals made by prediction errors in the experiments. This is why the length distribution did not improve the CRF-based methods.

5.2 Analysis of False Positives and False Negatives

So far, the discussion has shown that the proposed method performs well and is effective in stress prediction. However, there are still some false positives and negatives. False positives and negatives often occur when stressed words are determined by a certain intention of the chant text. Take for example the sentences *What does he want? He wants one egg*. In the standard manner, the words *What*, (the first) *want*, *He*, and *egg* get stressed. This is exactly what the proposed method did in the experiments. However, one could put stress on the word *one* instead of the word *He* to intend that he wants only ONE egg. The proposed method hardly handle the intention of a give chant text. It is indeed difficult to understand and deal with the intention by existing techniques. Considering this, it would be a better strategy that we first apply the proposed method to obtain the basic stress prediction results and then let teachers modify them according to the intention.

In addition to reducing false positives and negatives, the proposed method needs to be improved in another area. In the experiments, we assumed that null stressed words were given. In the real application, however, one needs to predict null stressed words in some cases. A simple idea for solving this problem is that if the prediction result does not satisfy the constraint on the number of stressed word, we can add some null stressed words to the prediction result and evaluate whether the probability improves or not. This will be our feature work.

6. Conclusions

In this paper, we described a method for stress prediction for automatically predicting stressed words in chant texts. We proposed exploiting several sources of information which are relevant to stress prediction by using CRFs. We also proposed methods for satisfying the constraint on the number of stressed words in a chant text and for considering the distribution of the length of stress-intervals. The experiments showed that the proposed method achieved an F-measure of 0.939 and outperformed the other methods implemented for comparison. The proposed method is expected to be useful in supporting non-native teachers of English when they teach chants to students and create chant texts with stress marks from arbitrary texts.

In future work, we will explore methods for generating null stressed words. We will also explore how we can apply the proposed method to chanting robots that interactively teach English rhythm based on chants.

References

[1] C. Graham, Creating Chants and Songs, Oxford University Press, Oxford, 2006.
[2] M.L.G. Gregory and Y. Altun, "Using conditional random fields to predict pitch accents in conversational speech," Proc. 42nd Annual Meeting on Association for Computational Linguistics, pp.47–54, 2004.
[3] A. Margolis and M. Ostendorf, "Acoustic-based pitch-accent detection in speech: Dependence on word identity and insensitivity to variations in word usage," Proc. 2009 IEEE International Conference on Acoustics, Speech and Signal Processing, pp.4514–4516, April 2009.
[4] R. Nagata, T. Mizumoto, K. Funakoshi, and M. Nakano, "Toward a chanting robot for interactively teaching English to children," Proc. INTERSPEECH Satellite Workshop on Second Language Studies: Acquisition, Learning, Education and Technology, P2-13, Sept. 2010.
[5] J. Lafferty, M. Andrew, and F. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," Proc. International Conference on Machine Learning, pp.282–289, June 2001.
[6] I. Witten and E. Frank, Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations, Morgan Kaufmann Publishers, 2000.

Ryo Nagata graduated from the Department of Electrical Engineering, Meiji Univ. in 1999 and completed the doctoral program in information engineering at Mie Univ. in 2005 and became a research associate at Hyogo Univ. of Teacher Education. Since 2008, he has been an associate professor at Konan University. His research interests are language modeling, grammatical error detection and correction, and edumining (educational data mining). He is a member of the Association for Natural Language Processing.

Kotaro Funakoshi is with Honda Research Institute Co., Ltd. since 2006. He received the B.S. degree in 2000 from Tokyo Institute of Technology, the M.S. and the Dr. Eng. degrees from Tokyo Institute of Technology in 2002 and 2005, respectively. His research interests are natural language understanding/generation, spoken dialogue systems and conversational robots. He is a member of ACM SIGCHI, IPSJ, JSAI and ANLP.

Tatsuya Kitamura received B.S. in Engineering from Yamagata Univ. in 1992 and M.S. and Ph.D. in Information Science from Japan Advanced Institute of Science and Technology in 1994 and 1997, respectively. He was a research associate at Shizuoka Univ. from 1997 to 2002 and he was a researcher of ATR Human Information Science Laboratories from 2002 to 2007. Since 2007, he has been with the faculty of Science and Engineering, Konan Univ., where he is now a professor in the Faculty of Intelligence and Informatics. He is a member of the Acoustical Society of America and the Acoustical Society of Japan.

Mikio Nakano is a Principal Researcher at Honda Research Institute Japan Co., Ltd. (HRI-JP). He received his M.S. degree in Coordinated Sciences and Sc.D. degree in Information Science from the University of Tokyo, respectively in 1990 and 1998. From 1990 to 2004, he worked for Nippon Telegraph and Telephone Corporation. In 2004, he joined HRI-JP. His research interests include spoken dialogue systems, speech understanding, and conversational robots. He is a member of ACM, ACL, ISCA, JSAI, IPSJ, RSJ, and ANLP.