Study of Utterance Support System using Word Input based on Word Prediction

著者	Hamai Yuya, Mizumachi Mitsunori, Nakatoh Yoshihisa
journal or publication title	Journal of the Institute of Industrial Applications Engineers
volume	2
number	1
page range	13-18
year	2014-01-25
URL	http://hdl.handle.net/10228/00007829

doi: https://doi.org/10.12792/jiiae.2.13

Journal of the Institute of Industrial Applications Engineers Vol.2, No.1, pp.13–18, (2014.1.25)
DOI: 10.12792/JIIAE.2.13 ISSN:2187-8811

Paper

Study of Utterance Support System using Word Input based on Word Prediction

Yuya Hamai* Member, Mitsunori Mizumachi* Member
Yoshihisa Nakatoh* Member, Kenji Matsui† Member

(Received November 18, 2013, revised December 27, 2013)

Abstract: Speech is very important for us to communicate with others. However, because of speech handicaps, there are many people who feel it difficult. Thus, we set our main purpose to study new utterance support system working on an information terminal. To make this system more effective, we made word prediction and added fixed phrase dictionary and class 2-gram as additional functions. Fixed phrase dictionary includes phrases often used in daily life conversations and class 2-gram using co-occurrence frequencies of 9 parts of speech in Japanese. And by using class 2-gram, we made two additional functions. One is generating next candidates and another is changing frequency. We did three evaluation experiments - necessary numbers of input, input speed and measurement experiments of voice synthesis. And results shows that our system needs less numbers of input but slower than other methods and it is faster to do voice synthesis when user input a sentence than other timing. Hereafter, we intend to create more effective dictionaries, improve class 2-gram and develop new GUI to make our system more effective and faster.

Keywords: utterance support, word prediction, class 2-gram, voice synthesis.

1. Introduction

Speech is very important for us to communicate with others in our daily lives. However, because of speech handicaps, some people find it difficult. The main approaches for the people are writing and sign language. But they are not so easy to master and writing needs to carry pens and papers at all times[1]. Nowadays, many new utterance support methods are studied[2]~[4]. For example, "GlovalVoice voice support" [5] is a communication tool working on pad, User inputs desired words then pad outputs them by using voice synthesis. And "Yubide Hanasou" [6] is an application that user can communicate with others by drawing strokes on his iPhone.

Recently, many people have their own information terminal, thus we set our main purpose is to study new utterance support system working on an information terminal (ex. Smartphone, pad). And to make this system faster and more effective than conventional methods, we studied word input methods and voice synthesis. Especially, we studied a word prediction.

Word prediction is one of word converting methods and it predicts user's desired words by using a small part of user's word input[7] [8]. To make word prediction faster and more effective, we used 1) fixed phrases dictionary and 2) class 2-gram as additional functions. Fixed phrases dictionary includes phrases often used in daily life conversations and class 2-gram is a contiguous sequence of 2 classes from a given sequence of text or speech. And we decided to use 9 parts of speech (POS) in Japanese as classes. And we did three evaluation experiments to find the efficient of both fixed phrases dictionary and class 2-gram.

In chapter 2, we will show the whole block diagram of this system. In chapter 3, our word prediction is explained. In chapter 4, we show three results of each evaluation experiments. In chapter 5, examination and in chapter 6, conclusion is given.

2. Whole Block Diagram of Proposal Utterance System

This system consists of three parts and Fig. 1 shows the whole block diagram. The input part processes the Roman character codes generated by user's keyboard inputs, and then outputs them to word prediction part. Information terminals like smartphones have their own GUI and user generally uses it to input words or phrases, but we used line Linux's input method and this might be more difficult. The

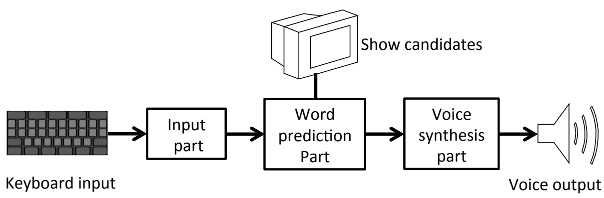


Figure 1: Whole block diagram.

* Affilate of First Author
1-1,Sensui-chou, Tobata-ku, Kitakyushu-shi, Fukuoka, Japan 804-0082
(n349427y@tobata.isc.kyutech.ac.jp)
† Osaka Institute of Technology.
5-16-1, Oomiya, Asahi-ku, Oosaka, Japan 535-0002

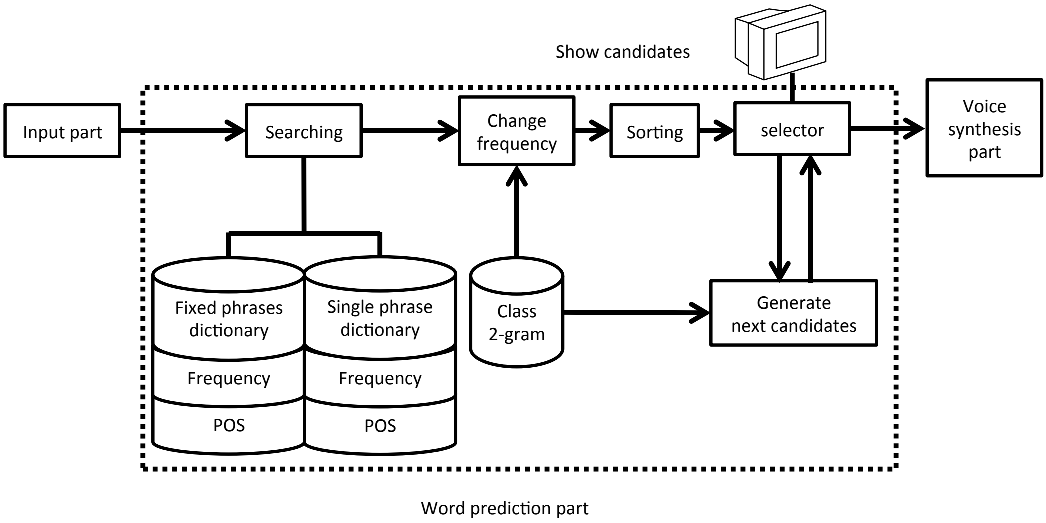


Figure 2: Proposed word prediction part.

word prediction part predicts words or phrases by using Roman character inputs processed in input part. The details of this part are shown in chapter 3. The voice synthesis part processes the words or phrases selected by user in word prediction part and output them as voice. Some kinds of voices will be available and user can choose one that he likes. However, in this paper, we used Open Jtalk(9) as a voice synthesis part.

3. Detail of Word Prediction Part

This chapter shows the details of the word prediction part. A word prediction is a word converting method that predicts the desired words or phrases by inputting small part of word. Currently, this method is used in many information terminals that contain smaller number of keys. Figure 1 shows the detail of this part. User's character inputs are used to search for words or phrases in dictionaries. The dictionaries have not only words or phrases but also POS and word frequencies which are used to sort and predict candidate words or phrases. In the sorting part, the candidates are sorted according to word and phrase frequencies and shown to user. The class 2-gram is used to change frequency and generate next candidates. In the selector part, sorted candidates and generated next candidates are showed to user. The details of 1) the dictionaries, 2) the word and phrase frequency and 3) the class 2-gram are shown in following sections.

3.1 Dictionaries There are two dictionaries in our proposed word prediction - a single phrase and fixed phrases - and user's input directly searches to both dictionary to make candidate words or phrases. The details of both dictionaries are shown in following.

3.1.1 Single Phrase Dictionary SKK dictionary (M size) [10] are mainly used to make this dictionary. The SKK dictionary contains many words of the noun but other POS are not so much, thus a morphological analysis is applied to Japanese example sentences on English conversation learning site using Chasen[11] and the results are added to the dictionary. Chasen parses sentences into words and analyzes it in each POS. After these operations, total number of words in this dictionary became 13,798.

The single phrase dictionary consists of 1) number, 2) pronunciation, 3) resulting word and 4) POS. Table 1 shows a concrete example of this dictionary. Pronunciation is used to search for words and all of them are written in Roman characters. The reason why Roman characters were used instead of Japanese is Roman characters input can show candidates in smaller inputs and can correct wrong spelling easily than using Japanese. These pronunciations were made by using kakasi[12] and all of them were written in Kunrei-shiki. Kunrei-shiki is a slightly modified version of Nihon-shiki which eliminates differences between the kana syllabary and modern pronunciation.

Resulting words were shown to user. POS is used to change word frequency in dictionary and show next candidate words. The details of POS usage are shown in chapter 3.3

3.1.2 Fixed Phrases Dictionary Fixed phrases dictionary consist of daily life conversation sentences as shown

Table 1: Example of single phrase dictionary.

Number	Pronunciations	Result word	POS
5992	katamuki	傾き	Noun
5993	katamusubi	片結び	Noun
5994	katame	片目	Noun
⋮	⋮	⋮	⋮

Table 2: Example of fixed phrases dictionary.

Number	Pronunciations	Result word	POS
12	gomennasai	ごめんなさい	Interjection
13	suimasen	すいません	interjection
14	mousiwakenai	申し訳ない	Noun
⋮	⋮	⋮	⋮

Table 3: POS in Japanese.

Sort	Independent words	Ancillary words
term	Verb Adjective Na-adjective	Noun
predicate	Noun Adnominal adjective Adverb Conjunction Interjection	Postpositional particle

in Table 2. This dictionary was made to decrease user's inputs.

This dictionary contains 550 phrases which were obtained from Japanese example sentences in Japanese learning sites for non-Japanese speakers. The structure of this dictionary is same as single phrase dictionary

3.2 Word and Phrase Frequency Word and phrase frequency changes its own values to learn user's habit.

All of words and phrases have their own frequency. When a word or phrase was selected by user, the frequency which has same number with it increases 1.

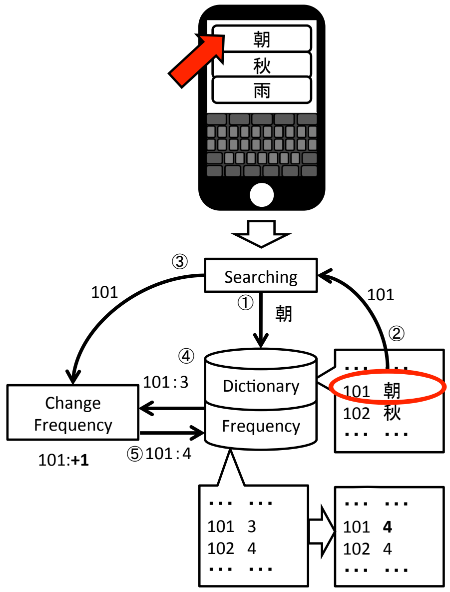
If a desired word was not found in sorted candidates, user chose nothing from it and all of frequencies of words and phrases in candidates decrease by half. After this operation, even user inputs same character as latest input, the probability of user's desired word or phrase coming up is increased. Figure 1 shows the example of these operations. Frequencies are also changed by an additional function of class 2-gram.

3.3 Class 2-gram Class 2-gram is a kind of n-gram. An n-gram is a contiguous sequence of n items from a given sequence of text or speech. The n-grams typically are collected from a text or speech corpus. An n-gram of size 2 is a "bigram"[13].

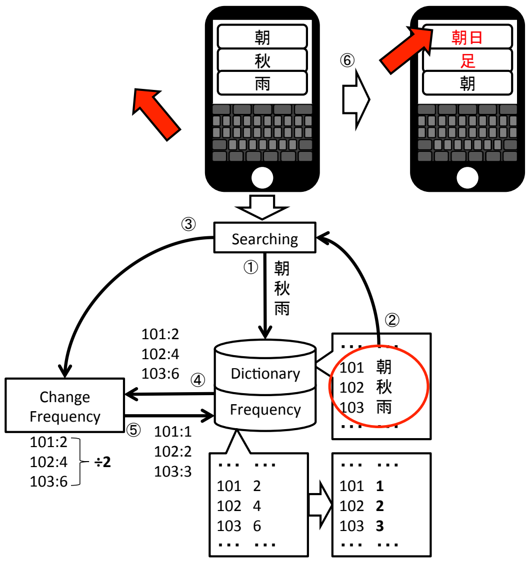
Class 2-gram is a contiguous sequence of 2 classes from a given sequence of text or speech. And 9 POS in Japanese are used as classes[14]. The reason why we used class 2-gram in this word prediction is that it has less data than 2-gram. For example, if there are 10,000 words in a word prediction dictionary, the amount of data is 10,000×10,000. However, class 2-gram has no connection with a number of words in dictionary and the amount of data is only 9×9. There are 10 POS in Japanese and table 4 shows them. Because some people didn't recognize na-adjective as a POS in Japanese, thus we use nine POS.

Figure 4 shows how a class 2-gram is made. We made it by doing morphological analysis to 4000 Japanese example conversation texts, picking up POS and analyzing their co-occurrence. Table 5 shows an example. By using this class 2-gram, two functions are addicted and they help word prediction to predict words and phrases more effectively. One is generating next candidates and another is Changing Frequency using class 2-gram.

3.3.1 Next Candidates The next candidates are automatically predicted by using class 2-gram. The words or phrases contains POS which are highest two of the class 2-gram and are automatically selected from dictionaries,



(a) User selects "morning"



(b) User selects nothing

Figure 3: Changing frequency.

Table 4: Example of class 2-gram.

Following POS	Noun	Verb	Adjective
	Probability[%]		
Conjunction	0.061318	0.117256	0
Noun	19.56055	13.30858	32.83396
Verb	7.235565	7.777995	16.22971
Modal verbs	11.17016	46.82431	27.8402
Interjection	0.132856	0.175884	0
Postpositional particle	60.98109	30.03713	20.09987
Adverb	0.224834	0.390854	0.249688
Prefix	0.194175	0.234512	0.249688
Adjective	0.255493	0.605824	1.498127
Adnominal adjective	0.183955	0.527653	0.998752

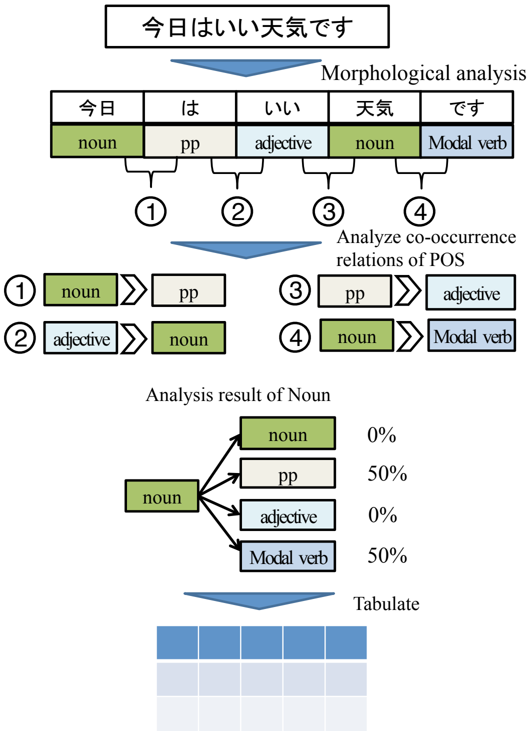


Figure 4: Way of making class 2-gram.

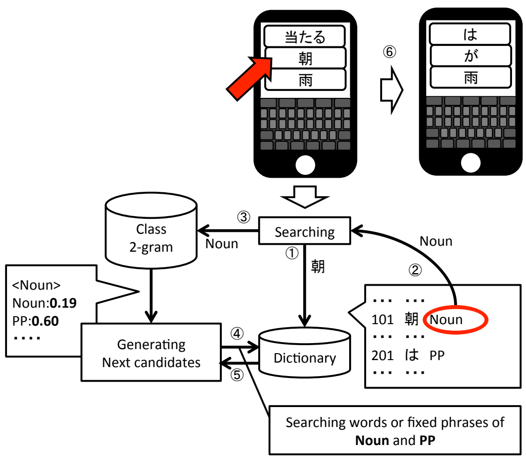


Figure 5: Way of generating next candidates.

sorted according to their frequencies and shown to users as next candidates. For example, if the user chooses a noun, next candidates are particle or noun. Figure 5 shows the way of generating next candidates

3.3.2 Changing Frequency by Using Class 2-gram

By using the latest result of the word prediction, word and phrase frequencies are changed by multiplying probabilities of the class 2-gram. For example, if the latest result was noun, then word and phrase frequencies of particle are multiplied by "0.6" and so are others. After this operation, all of POS except particle and noun become difficult to show up. Figure 6 shows the concrete example.

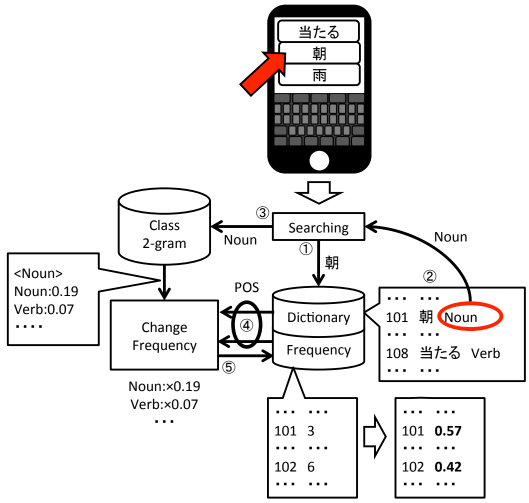


Figure 6: Way of changing frequency by using class 2-gram.

Table 5: Number of subjects.

Evaluation	Number of subjects
Necessary numbers of input	1(author)
Input speed	5(including author)
measurement experiments of voice synthesis	1(author)

Table 6: Result of number of input.

Methods	Numbers of input (/100 characters)
proposed word prediction	119
proposed word prediction +class 2-gram	102
iPhone	171
Microsoft Office IME	221

4. Experiments

We evaluated our system through three experiments. And we mainly tried to find out our word prediction's performance by adding additional functions and measurement experiments of voice synthesis.

4.1 Methods

In this paper, we three evaluation experiments - necessary numbers of input (/100 characters), input speed (/1 min) and measurement experiments of voice synthesis - and each result is shown in following sections. This study especially focused on the efficiency of class 2-gram. Sentences used in necessary numbers of input and input speed experiences are from "86th Examination in Japanese Word Processing" and four sentences that is supposed daily life conversations is used in measurement experiments of voice synthesis. We compare our word prediction with the Microsoft Office IME 2010 (MS IME) and the iPhone. We used Wired Keyboard 600 to input character to the word prediction and MSIME. Table 6 shows the number of subjects of each experiment.

4.2 Results

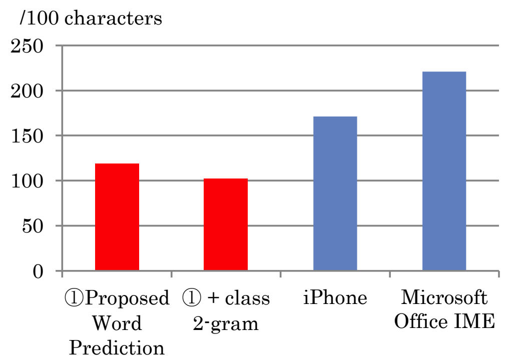


Figure 7: Result of number of input.

Table 7: Result of input speed.

Methods	Input speed(/min)				
	subjects				
	1	2	3	4	author
proposed word prediction	41	29	42	41	81
proposed word prediction +class 2-gram	35	33	34	57	85
iPhone	84	65	79	67	77
Microsoft Office IME	101	57	93	76	98

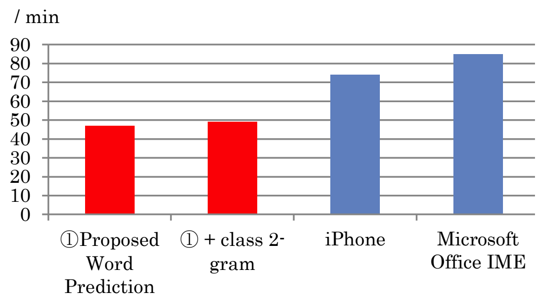


Figure 8: Result of input speed.

4.2.1 Necessary Numbers of Input

Table 6 and Fig. 7 show the experimental result of necessary numbers of input. There are two results of our proposed system - one is itself and other is with class 2-gram. According to Table 7, proposed word prediction needed less numbers or input than other methods. And adding class 2-gram makes it more effective with a 40% lead over the iPhone.

4.2.2 Input speed

Table 7 shows the result of input speed and Fig. 8 shows the mean value of each subject. The values indicate how many characters subjects could enter and faster methods have bigger number. The results from the subjects show that using our word prediction is slower than other methods. Even though we added class 2-gram; the iPhone was still twice as fast as ours and some results become slower.

4.2.3 Measurement experiments of voice synthesis

Table 8 shows the result of measurement experiments of voice synthesis. We did two measurements about two timings of voice synthesis. One is when subject input a word or phrase and another is when he input a sentence. The result shows that doing voice synthesis when user input a sentence is faster.

Table 8: Result of measurement experiments of voice synthesis.

	Word or phrase[s]	sentence[s]
Voice synthesis	68.0	24.4
Total	122	99.1

5. Examination

The necessary numbers of input became lower than other methods by using our word prediction. We conjectured that using two different dictionaries and class 2-gram had a good effect on ours. According to the result, using next candidates and frequency change especially made our system 10% faster. From this, using a class 2-gram is effective to add to word prediction.

The result of input speed was worse than other methods. The questionnaires we send out after the experiments showed that Linux line input makes our proposed word prediction hard to see candidates. In actuality, only my result was better than others because I am familiar with it. It is conceivable that Kunrei-shiki Roman character input which was the trait of our word prediction made it difficult to input. According to the questionnaires, Kunrei-shiki is unfamiliar and it was difficult to confirm entered words.

Measurement experiments of voice synthesis show that doing voice synthesis when user input a sentence is faster than doing when he input a word or phrase. This may be due to the processing time of voice synthesis part. Conversely, processing when user input a word or phrase is faster about the timing that user began to speak. Thus now we are considering a new timing of voice synthesis which can start speaking faster and needs short time to end speaking.

6. Conclusion

We study new utterance support system working on an information terminal for the people who have a speech handicap. We used word prediction and added class 2-gram to this system. The experience results show that our system needs less numbers of input but slower than other methods and it is faster to do voice synthesis when user input a sentence than other timing. Hereafter, we intend to create more effective dictionaries, improve class 2-gram and develop new GUI to make our system more effective and faster.

Acknowledgment

This work was partially supported by a Grant-in-Aid for Science Research (No.24500664) from the Japan Society for the Promotion of Science.

References

[1] Ministry of Health, Labour and Welfare, "Survey on persons with physical disability", pp.4-25, 2006.

[2] K. Yabu, T. Ifukube, "A speech synthesis device controlled by traced position on a touch control surface -singing voice production based on a melody data-", *IEICE General Conference 2013*, No.A-19-11, pp.272, 2013.

[3] H. Tanaka, Y. Nagashima, "Study of Japanese Sign Language Semantic Process based on N400 Analysis", *IEEE International Conference on Systems, Man and Cybernetics, 2006. SMC '06.*, pp.353-357,2006. DOI:10.1109/ICSMC.2006.384407

[4] F. Fang, T. Shinozaki, "Continuous electrooculogram recognition for an eye-motion-based communication aid system", *2012 Autumn Meeting, Acoustical Society of Japan*, No.1-4-4, pp.1513-1514, 2012

[5] Glovalvoice, "http://voicetext.jp/gv/", *HOYA Service Corporation*, access date: 2013.10.09

[6] Yubi de Hanasou, "http://www.dencom.co.jp/yubihana/ yubihana.htm", *Densei Communication Inc.*, access date: 2013.10.09

[7] T. Masui, "POBox: An Efficient Text Input Method for Handheld and Ubiquitous Computers", *Proceedings of the International Symposium on Handheld and Ubiquitous Computing - HUC '99*, pp.289-300, 1999.

[8] K. Tsugioka, H Tahashi, T Yoshikawa and T. Furuhashi, "A Study on Implementation of Predictive Text to P300 Speller for Improving Spelling Speed". *Journal of Japan Society for Fuzzy Theory and Intelligent Informatics*, Vol.24, No.1, pp.553-559, 2012

[9] Open JTalk, "http://open-jtalk.sourceforge.net/", *Open JTalk SourceForge page*, access date: 2013.10.09

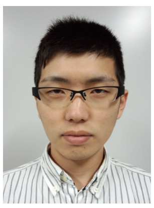
[10] SKK Openlab, "http://openlab.ring.gr.jp/skk/dic-ja.html", access date: 2013.10.09

[11] Morphological Analysis Software, "http://chasen-legacy.sourceforge.jp/", access date: 2013.10.09

[12] Kanji Kana Simple Inverter, "http://kakasi.namazu.org/", access date: 2013.10.09

[13] G. W. Lesher, B. J. Moulton and D. J. Higginbotham, "Effects of Ngram Order and Training Text Size on Word Prediction", *http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.23.7518&rep=rep1&type=pdf*, 1999.

[14] T. Shimizu, Y. Hirofumi, M. Shoichi and S. Yoshinori, "Spontaneous Dialogue Speech Recognition using Cross-Word context Constrained Word Graph", *IEICE technical report. Natural language understanding and models of communication*, Vol.95, No.120, pp.49-54, 1995.



Yuya Hamai (Member) He was born in Hiroshima, Japan, on December 1, 1990. He received a B.E. degree in engineers from Kyushu Institute of Technology in 2013, and is presently a master course student at Graduate School of Engineering, Kyushu Institute of Technology. He is member of IEICE and ASJ.

Mitsunori Mizumachi (Member) He received a Bachelor degree in design from Kyushu Institute of Design and a Ph.D degree in information science from Japan Advanced Institute of Science and Technology (JAIST) in 1995 and 2000, respectively. He is currently an associate professor at Kyushu Institute of technology. His research interests include acoustic information processing and statistical signal processing. He is a member of AES, ASA, ASJ, IEEE, IEICE, and RISP.

Yoshihisa Nakatoh (Member) received his Ph.D degree from shinshu University in 2007. He worked for Sharp Corporation from 1986 to 1989 and worked for Matsushita Electric Industrial Co. Ltd from 1991 to 2008. He is currently a professor at Kyushu Institute of technology. His research interests include speech recognition, audio coding, hearing aid and accessibility technology. He is a member of ASJ, IEICE, IPSJ.

Kenji Matsui (Member) received a Ph.D. degree in electronics and information engineering from Osaka City University in 2002, and is presently a professor at Osaka Institute of Technology. He has worked on speech signal processing. He is member of IIAE, IEICE, and ASJ.