Construction and Analysis of Corpus of Japanese Classroom Lecture Speech Contents

Masatoshi Tsuchiya* and Satoru Kogure† and Hiromitsu Nishizaki‡ and
Kazumasa Yamamoto§ and Seiichi Nakagawa§
* Information and Media Center, Toyohashi University of Technology, Toyohashi 441–8580 Aichi Japan
E-mail: tsuchiya@imc.tut.ac.jp Tel: +81-532-44-1308
† Faculty of Informatics, Shizuoka University, Hamamatsu 432–8011 Shizuoka Japan
E-mail: kogure@inf.shizuoka.ac.jp Tel/Fax: +81-53-478-1477
‡ Department of Research Interdisciplinary Graduate School of Medicine and Engineering, University of Yamanashi,
Kofu 400–8511 Yamanashi Japan E-mail: hnishi@yamanashi.ac.jp Tel: +81-55-220-8483
§ Department of Information and Computer Sciences, Toyohashi University of Technology, Toyohashi 441–8580 Aichi Japan
E-mail: {kyama,nakagawa}@slp.ics.tut.ac.jp Tel: +81-532-44-6777

***Abstract*—This paper explains our developing *Corpus of Japanese classroom Lecture speech Contents* (henceforth, denoted as CJLC). Increasing e-Learning contents demand a sophisticated interactive browsing system for themselves, however, existing tools do not satisfy such a requirement. Many researches including large vocabulary continuous speech recognition and extraction of important sentences against lecture contents are necessary in order to realize the above system. CJLC is designed as their fundamental basis, and consists of speech, transcriptions, and slides that were collected in real university classroom lectures.**

I. Introduction

Recently, there is increasing interest in interactive e-Learning systems like exCampus[1], IT's class[2] and Blackboard,[3] because they enable students to learn anywhere and anywhen they want. All of these systems, however, share a big fault: they can treat texts of slides, but not speech. It means that users can not search slides with keywords which are uttered in the speech, although they can search slides with keywords which occur in titles and texts of slides. In order to realize an interactive e-Learning system which can treat both texts and speech, several technologies like spoken document retrieval[11], video content analysis[8] and automatic speech summarisation[5][14] are necessary. Especially, robust speech recognition of lecture speech is the most important technology among them.

There are, however, various problems with recognising real classroom lecture speech: speaking styles of teachers, influence of microphones used when recording their speech, noise and/or reverberation of classrooms and language models which cover lecture-related contents. A corpus of classroom lecture speech which is designed particularly for these problems is obviously required, in order to cope with these problems.

We already have several corpora of classroom lecture speech. The MIT research group[13][4] has created a corpus, including more than 300 hours of English classroom lectures from eight different courses and 80 seminars given on a variety of topics at MIT. This corpus is, however, insufficient to evaluate influence of a generally used lapel microphone, because these data were recorded with an omni-directional microphone under a general classroom environment. LECTRA[9][15], which is the national project in Portugal, includes total 23 Portuguese lectures (approximately 5.2 hours and 44k words included) from two different courses recorded with a lapel and head-mounted microphones. This project also reported the performance of recognising lecture speech and analysed recognition errors.

Corpora of general spontaneous speech are possible resources to solve the described problems. The Rich Transcription (RT) evaluation series [4] that have started since 2002 are implemented to promote and gauge advances in the state-of-the-art in several automatic speech recognition technologies by using spontaneous speech[2][1]. In the recent RT evaluation, the tasks of "Speech to Text" (STT), "Speaker Diarization", and "Speech Activity Detection" (SAD) have been evaluated on the three meeting domains. Corpus of Spontaneous Japanese[10] (henceforth, denoted as CSJ) is the biggest corpus of spontaneous Japanese including about 1,000 academic presentations and about 1,600 simulated public speeches. As each speech included in CSJ was recorded with a headset microphone, it is impossible to use CSJ for evaluating influences caused by microphone types. Furthermore, a corpus of classroom lecture speech is still required, because there is the difference in disfluency acts between academic presentation speech and classroom lecture speech as described later in Section III-A.

This paper explains our ongoing project called as *Corpus of Japanese classroom Lecture speech Contents* (CJLC). CJLC is designed as a fundamental basis for developing technologies of robust speech recognition and advanced processing of e-

[1] http://excampus.nime.ac.jp/index.html
[2] http://www.gp.hitachi.co.jp/eigyo/product/itsclass/
[3] http://www.blackboard.com/us/index.Bb
[4] http://nist.gov/speech/tests/rt/index.htm

Learning contents, and consists of a lot of Japanese classroom lecture speech recorded at several universities. Furthermore, we are going to release CJLC publicly for research usage. We hope that CJLC makes a breakthrough in the technologies of spoken language processing for e-Learning contents.

Reminder of this paper is organised as follows: Section II describes the detailed specification of CJLC, and Section III presentes difference between classroom lecture speech and academic presentation speech, and influences of microphone performances and language models for LVCSR performance. We conclude in Section IV.

II. Specification of CJLC

As described before, there are several problems that impair recognition of classroom lecture speech. CJLC is especially designed to resolve two problems among them. The first one is to evaluate influences caused by microphone types under noise and reverberation environment of real classrooms. Speech of CJLC, therefore, are recorded in real classrooms with several microphones. The second one is various speaking styles and widespread lecture topics. CJLC covers many speakers at several universities to evaluate influences caused by speaking styles, and consists of many courses at computer science departments such as physics, electronics, mathematics and information sciences. The rest of this section explains the detailed specification of CJLC.

A. Structure of CJLC

CJLC is formally defined as a set of classroom lecture data, and each dataum consists of the following items:

- a lecture speech recorded with several microphones,
- its synchronised transcription,
- a presentation slide data (optional, Microsoft PowerPoint format),
- a timetable for the slide show (optional), and
- a list of important utterances (optional).

A lecture speech dataum and its synchronised transcriptions are provided for all lectures, but a presentation slide dataum, a timetable of slide show and a list of important utterances are attached to not all lectures. EZ presentator which is an e-Learning software made by Hitachi Advanced Digital Inc. is used to record a timetable of the slide show.

Table I shows the statistics of CJLC. Because each speaker lectures one or more courses, the number of speakers is less than the number of courses. Furthermore, several lectures are recorded for each course as shown in Table I. 6 lectures among the total 89 lectures contain lists of important utterances, which are annotated by 6 professional researchers.

Fig. 1 shows the distribution of CJLC lecture durations. It is notable that CJLC lectures can be classified into two categories: the lectures which are shorter than 60 minutes and the lectures which are longer than 60 minutes.

B. Recording Condition of CJLC

A lapel microphone is widely used instead of a hand held microphone when recording lectures, because it frees the teachers' hands and does not prevent a lecture, although it drops recognition accuracy generally. This means that it is important to investigate performance difference among microphone types on speech recognition and to find a compensation method. A speech data of CJLC, therefore, contains multi-channel data recorded with both a lapel microphone and an other type microphone, unlike previous corpora. For example, a speech datum of CJLC contains the data recorded with a lapel microphone and the data recorded with a headset microphone. Table II shows various recording conditions for recording speech of CJLC. And more, the speech was recorded at classrooms without special audio equipment, in order to make a corpus under noise and reverberation environment of real classrooms.

TABLE I
Statistics of CJLC

# of speakers	15
# of courses	26
# of lectures	89
Duration	3,780 min.

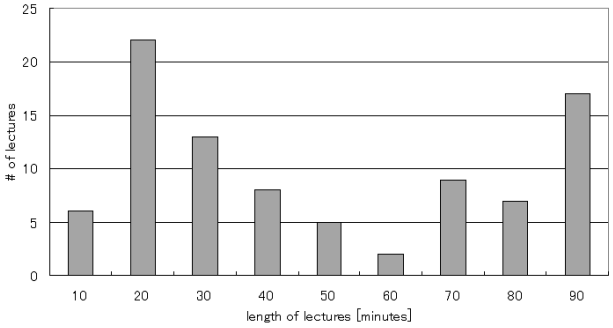


Fig. 1. Number of Lectures and Durations

C. Transcription Format of CJLC

To provide data for acoustic and language model training, we created manual transcriptions of the lecture speech. Each speech was automatically segmented into utterances using the power information of speech described in [6], [12]. The annotators were instructed to pay careful attention to generate a transcription of what was spoken. They were also instructed to annotate speech phenomenas in utterances with the following

TABLE II
Recording Conditions of CJLC

Microphone type	Recording hardware	Format of speech
wireless lapel (TOA WM-1300)	DAT recorder (Sony TCD-D8)	48KHz 16bit PCM
wired hand held (Sony C-355)		
wired headset (Shure SM10A)	IC recorder (Marantz PMD-671)	16kHz 16bit, PCM

TABLE III
DIFFERENCE BETWEEN CJLC AND CSJ

	CJLC	CSJ
main target	classroom lectures	lectures in a academic meeting
# of lectures	89	3300
microphone	various microphones (described before)	headset (CROWN CM-311A)
duration per a lecture	long	short
annotation (tagging)	11 tags (CSJ sub-set)	CSJ tags
transcriptions of speech	YES	YES
Slide data	YES (partially)	NO

:
0147: でこちらは (F えーと) 発展課題，ですね (F えーと)
0148: やりたい方，(F えーと)
0149: 興味がある (D の) 方はやってみてください
0150: (F えーと) さっきのところって言うのは，データの性質に関わらず，(A エヌ; N) の値のみで決まるデータがどこかというのを
0151: やりました
:
(translated into English)
0147: And here (F well) is the extended exercises, (F well).
0148: The person who want to exercise it, (F well).
0149: Please try (D a) it if you are interested.
0150: (F Well) what I said a little while ago is where is the data decided only by the value of (A enu;N) without the characteristic of data...
0151: I've taught it.
:

Fig. 2. An example of transcription

11 kinds of tags:

- (F) filled pauses,
- (D) fragment of content words,
- (D2) fragment of functional words,
- (A) numerical and alphabetical representation,
- (W) corruption,
- (L) spoken word(s) with a laugh,
- (T) blubbered spoken word(s),
- (C) spoken word(s) with cough,
- <C> a sound of cough,
- <B> a sound of breath,
- <N> a noise, and
- <V> bubble of voices

These tags are a sub-set of CSJ tag set described in [7], and are compatible to CSJ because CSJ tagging policy is employed when annotating these tags. Tag **(F)**, Tag **(D)** and Tag **(D2)** are especially important to investigate disfluency acts in lecture speech.

Fig. 2 shows an example of transcription of CJLC. Each line, which is corresponding to an utterance unit, consists of two columns: the first column denotes the utterance sequential number in the whole lecture, and the second column shows the transcription of the utterance. In Fig. 2, the Japanese word "えーと", which means "well" in English, is annotated by Tag **(F)** as a filled pause. Tag **(D)** is also employed to annotate the word fragment "の" as a disfluency act.

TABLE IV
STATISTICS OF CJLC AND CSJ

		CJLC lectures	CSJ APS	CSJ SPS	CSJ dialogue
# of lectures		89	987	1715	58
# of words / lec.		6636	3358	2122	2613
duration / lec. [sec]		2610	1003	694	765
total duration [hours]		63.0	275.0	330.6	12.3
# of tags / lec.	F	410.3	229.2	118.8	322.2
	D	49.9	44.5	26.0	43.9
	D2	3.9	3.4	1.4	1.4

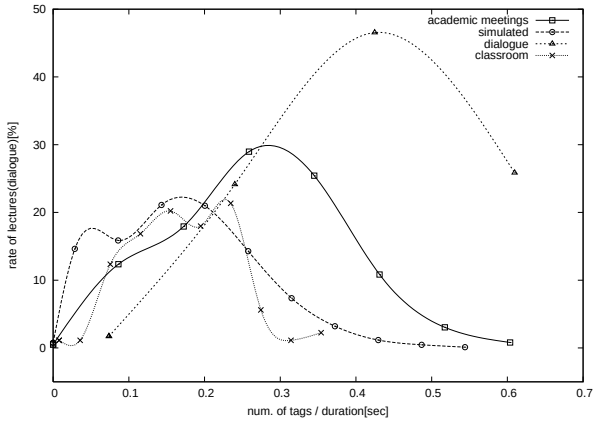


Fig. 3. Distribution of Tag (F) of CJLC and CSJ

III. ANALYSIS OF CJLC

A. Comparison of CJLC and CSJ

This section explains the difference between classrooms lecture speech and presentation speech.

CSJ is the biggest corpus of spontaneous Japanese, and contains four categories of speech: *academic presentation speech* (APS), *simulated public speech* (SPS), dialogue, and reading. APS is the live recording of academic presentation in 9 different academic societies covering the fields of engineering, social science, and humanities. SPS, on the other hand, is studio recording of layman speaker's speech of about 10–12 minutes, on everyday topics like 'the most delightful/saddest

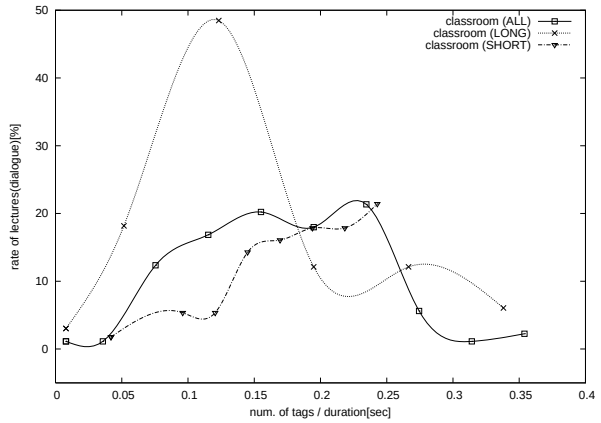


Fig. 4. Distribution of Tag (F) of CJLC Short/Long Lectures

memory of my life'. Table III summarises the differences between CJLC and CSJ, which are target speech, transcription tags of phenomenon in spontaneous speech, and slide data.

We have compared the phenomenon in spontaneous speech included in CJLC with that of CSJ. We especially analysed the frequency of filled pauses and disfluently spoken words in each corpus. Table IV represents the detailed statistics of two corpora. The numbers in Table IV mean the average values per lecture (or dialogue) for each item except the number of lectures and the total duration.

Fig. 3 shows frequency distributions of filled pauses annotated by Tag F at classroom lectures, APS, SPS and dialogues. As shown in Fig. 3, more filled pauses occur in the dialogue speech of CSJ than in the other types of speech. Classroom lectures of CJLC, APS and SPS share similar frequency distributions of filled pauses. Although Fig. 4 suggests that filled pauses occur more frequently in short lectures than in long ones, we think that there is no serious distinction among them, because both the distribution of filled pauses on short lectures and the one on long ones are still similar to the one on SPS as shown in Fig. 3.

B. Experimental Setup of LVCSR Performance on CJLC

We investigated how the differences between microphones affect the performance of speech recognition. Classroom lecture speech was recorded with four types of microphones. We also investigated the differences in recognition performances between language models. We used seven different language models; three types of language models trained using CSJ, two language models created using news articles and two models using Web collections for the experiments.

Table V shows the details of the lecture speech which was selected from CJLC. Various lectures from five courses were prepared.

We used Julius rev.3.5.3[5], for speech recognition, which is an open source decoder for LVCSR and runs in two decoding passes; the first pass uses a word bigram and the second pass uses a word trigram. Based on the word trigram and context-dependent HMM, Julius can perform real-time decoding on a 20k vocabulary dictation task in most of the current PCs.

[5]http://julius.sourceforge.jp/

TABLE VI
TRAINING CONDITIONS OF LANGUAGE MODELS.

LM ID	Vocab.	Training Data	
		# of Lectures	Size [Byte]
$CSJ_{970-20k}$	20k	970 *1	23 MB
$CSJ_{3300-20k}$	20k	3285 *2	123 MB
$CSJ_{3300-40k}$	40k		
$NEWS_{20k}$	20k	1,499,936*3	1400 MB
$NEWS_{42k}$	42k		
WEB_{20k}	20k	—*4	100 GB
WEB_{60k}	60k		

*1: 970 APS lectures

*2: 3285 CSJ lectures

*3: all articles from 1991 to 2004 on Mainichi newspaper, Japan.

*4: The LM was prepared by Fujii *et al.* [3] and trained using the 100G text of Web collection.

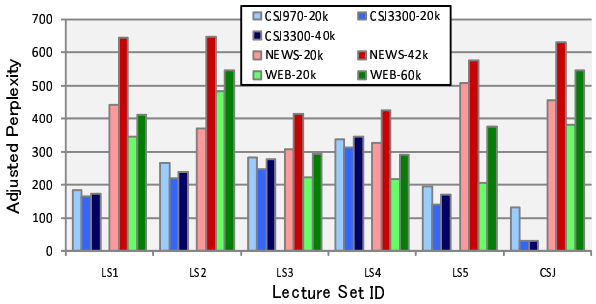


Fig. 5. Test set adjusted perplexity based on different language models.

Julius used triphone-based HMMs trained using CSJ recorded with a high-quality headset microphone (CROWN CM-311A), which were sampled at 16 KHz and 16 bits. Feature vectors comprised of 38 dimensions: 12 dimensional Mel-frequency cepstrum coefficients (MFCCs), the cepstrum difference coefficients (delta MFCCs), their acceleration (delta delta MFCCs), delta power, and delta delta power; these vectors were calculated every 10 ms. The distributions of the acoustic features were modeled using 32 mixtures of diagonal covariance Gussians for the HMMs.

Table VI shows the training conditions of each language model. Three types of models based on CSJ, which differed in the number of lectures of training and vocabulary size, were used in the decoder. Two types of models based on Mainichi newspaper articles, that differed only in vocabulary size, were used. Furthermore, two types of models that had a large vocabulary of 20k and 60k, respectively, were trained using Web articles.

C. Influence of Microphone Performance

In order to test the influence of microphone performance or wired/wireless condition, we recorded the two classroom lectures, L3 and L5 in Table V, with four types of microphones. The four microphones are (a) SONY ECM-C10 (normal/lapel), (b) SONY ECM-88B (high/lapel), (c) SONY

TABLE V
LECTURE SPEECH USED FOR RECOGNITION EXPERIMENTS.

Lec.ID	Spk.ID	#Snt	Time [s]	Filler Rate [%]	Course (Lecture Set ID)	Keyword for Lecture Contents
L1	S1	259	1213	6.80 (263/3869)	LS1: Computer Applications II	Spoken Language Processing
L2		236	1274	4.54 (179/3946)		DP Matching
L3		212	160	4.70 (80/1702)		Language Model
L4	S2	668	937	14.75 (496/3364)	LS2: Computer Applications I	Natural Language Processing
L5		311	254	10.09 (187/1853)		
L6	S3	1480	3623	6.64 (1006/15157)	LS3: Software Engineering	Design of Program, coding
L7	S4	743	1903	8.20 (484/5901)	LS4: Experiments on Physics	Diode, P-type and N-type Semiconductor
L8	S5	1163	4193	5.69 (672.11803)	LS5: Algorithm and Data Structure I and Practice	Binary Tree II
L9		903	3115	6.57 (550/8367)		Bubble Sort
L10		820	3285	8.24 (745/9037)		Selection and Insertion Sort
L11		564	2261	7.72 (504/6529)		Quick Sort
CSJ		1771	4568	10.91 (2050/18793)	CSJ	The Acoustical Society of Japan, etc.

TABLE VII
WORD RECOGNITION RATES OF LECTURES RECORDED WITH FOUR MICROPHONES[%].

(AM: Triphone / LM: $CSJ_{3300-40k}$)

mic.	(a) normal lapel	(b) high lapel	(c) normal handheld	(d) normal headset
Acc.[%]	55.4	56.4	60.0	62.7
Cor.[%]	61.3	62.7	67.3	70.5

ECM-355 (normal/handheld), and (d) ISOMAX Headset Microphone (normal/headset).

In this experiment, $CSJ_{3300-40k}$, as represented in Table VI, was used for speech recognition. Table VII shows the results of the experiment on the average of L3 and L5. From these, we could state the order of recognition performance as follows: headset microphone $>$ handheld microphone $>$ lapel microphone (high performance) $>$ lapel microphone (normal performance). In order to get a higher recognition rate for lecture speech, a higher-quality microphone should be used to record speech.

D. Language models

As described in Table VI, we prepared the seven language models and evaluated them on test set perplexity and speech recognition performance.

We calculated the *test set adjusted perplexity* (APP), which was given by adjusting the PP for taking account of OOV words [16], for five lecture sets (from LS1 to LS5) and the CSJ test set given in Table V. The values of APP and rate of OOVs are shown in Fig. 5 and Fig. 6. For almost all of the lectures, the values of APP of the CSJ_{3000} model were lower than the ones of the CSJ_{970} model. This was because the utterances in usual lectures that often include short dialogue, discourse, or monology were similar to the contents of the 3,300 lectures in CSJ, whereas all the training data of the 970 lectures were based only on the lectures presented at academic meetings in Japan. Moreover, the values of APP were high in the order CSJ $>$ Web $>$ NEWS because the news articles consisted of formal sentences and the sentence style in the Web collection was casual (similar to that of CSJ) but had formal sentence structures (similar to that of NEWS). Figure 7 shows the filler rate in OOVs. Four language models, WEB-

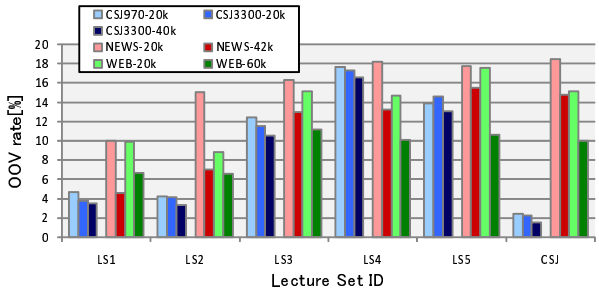


Fig. 6. Rate of OOVs based on different language models.

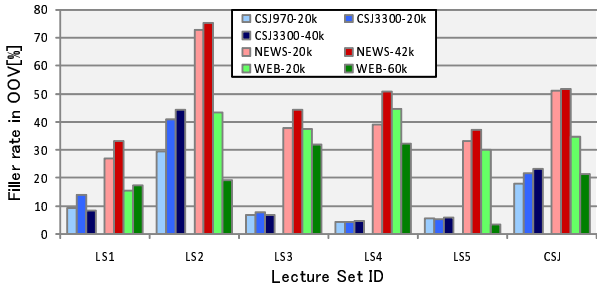
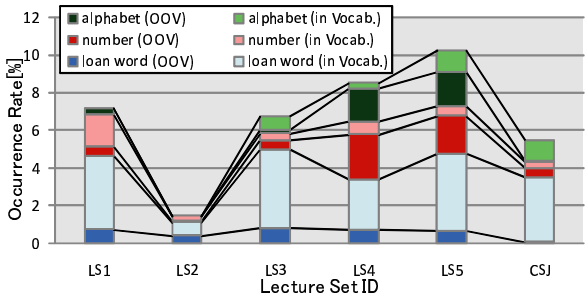


Fig. 7. Filler rate in OOVs of different language models.

20k, WEB-60k, NEWS-20k, and NEWS-42k contain at most several dozen filler words (that is, 9, 17, 2, and 4 types of filler words, respectively). On the other hand, three CSJ language models, 970-20k, 3300-20k, and 3300-40k contain about 1,000 types of filler words. Figure 8 shows the occurrence rate of alphabet, number and loan words. The rates of number (OOV) and alphabet (OOV) for LS4 and LS5 are higher than the ones of LS1, LS2, LS3, and CSJ because the contents of LS4 and LS5 are quite different from the contents of CSJ.

Next, we evaluated these language models on the basis of the speech recognition rate of five lecture sets and the CSJ test set. Figure 9 shows the word accuracy of the five lecture sets and the CSJ lecture set. The results showed that the lecture recognition using the $CSJ_{3300-20k}$ language model had a better performance than that using the CSJ 970 language model. In particular, the difference in the recognition performance was



LM: $CSJ_{3300-40k}$

Fig. 8. Occurrence rate of alphabet, number, and loan words

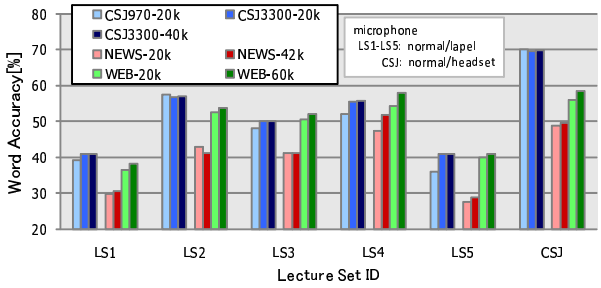


Fig. 9. Word accuracy of lecture speech recognition.

salient in LS3, LS4, and LS5. The word accuracy using WEB-20k was about the same as that of $CSJ_{3300-20k}$. Moreover, the recognition performances using NEWS language models were determinately bad when compared to the word accuracy of CSJ or Web collection language models. These results were attributed to the fact that the filler rate in the OOVs of NEWS was considerably higher than that in the OOVs of the CSJ or Web collection language models as shown in Fig. 7. Moreover, the filler rate of WEB-60k was lower than that of WEB-20k because the filler words that had a high frequency in lecture speech were contained in the vocabulary of the WEB-60k model but were not contained in the vocabulary of WEB-20k model.

From these results, we concluded that it was better to use the language models trained using a spontaneous or casual expression corpus than to use the language models trained using a formal/written style expression corpus for lecture speech recognition.

IV. Conclusions

This paper explains our developing corpus, called CJLC. Its main aim is to study the current state of classroom lecture speech recognition, one of the fundamental technologies needed to process the lecture contents. It consists of many real classroom lectures collected at a couple of universities that cover various lecture topics related to information sciences and cover various speaker types. And more, it is possible to evaluate influences caused by various microphones because they contain multi channel recorded speech.

Furthermore, we presented differences between classroom lecture speech and academic presentation speech, and influences of microphone performances and language models for the LVCSR performance.

The monitor version of CJLC is already available. Please see http://www.slp.ics.tut.ac.jp/CJLC/. We are going to release the formal version of CJLC database to the public limited to only research usage in near future.

Acknowledgment

This research was supported by Strategic Information and Communications R&D Promotion Programme of Ministry of Internal Affairs and Communications in Japan. Furthermore, we also thank teachers for their cooperation to record the classroom lecture speech.

References

[1] C. Fügen, M.Kolss, D. Bernreuther, M. Paulik, S. Stüker, S. Vogel, and A. Waibel. Open domain speech recognition & translation: Lectures and speeches. In *Proc. of ICASSP2006*, pages 569–572, 2006.

[2] Christian Fügen, Matthias Wölfel, John W.McDonough, Shajith Ikbal, Florian Kraft, Kornel Laskowski, Mari Ostendorf, Sebastian Stüker, and Kenichi Kumatani. Advances in lecture recognition: The isl rt-06s evaluation system. In *Proc. of Interspeech2006-ICSLP*, pages 1229–1232, 2006.

[3] A. Fujii and K. Itoh. Building a test collection for speech-driven web retrieval. In *Proc. of EUROSPEECH2003*, pages 1153–1156, 2003.

[4] James Glass, Timothy J. Hazen, Lee Hetherington, and Chao Wang. Analysis and processing of lecture audio data: Preliminary investigations. In *Proc. of the Workshop on Interdisciplinary Approaches to Speech Indexing and Retrieval at HLT-NAACL 2004*, pages 9–12, 2004.

[5] Chiori Hori, Takaaki Hori, and Sadaoki Furui. Evaluation method for automatic speech summarization. In *Proc. of EUROSPEECH2003*, pages 2825–2828, 2003.

[6] N. Kitaoka, T. Yamada, S. Tsuge, C. Miyajima, T. Nishiura, M.Nakayama, Y. Denda, M. Fujimoto, K. Yamamoto, T. Takiguchi, S.Kuroiwa, K. Takeda, and S. Nakamura. CENSREC-1-C: Development of evaluation framework for voice activity detection under noisy environment. In *IPSJ technical report, Spoken Language Processing (SIG-SLP), Vol.2006, No.107*, pages 1–6, 2006.

[7] H. Koiso, Y. Mabuchi, K. Nishizawa, M. Saito, and K. Maekawa. The specifications of transcriptions version 1.0. In *the Document of Corpus of Spontaneous Japanese*, pages –, 2003.

[8] Y. Li and C. Dorai. Instructional video content analysis using audio information. *IEEE Trans. Audio, Speech, and Language Process.*, 14(6):2264–2274, 2006.

[9] L.Lamel, E.Bilinski G. Adda, and J.L. Gauvain. Transcribing lectures and seminars. In *Proc. of EUROSPEECH2005*, pages 1657–1660, 2005.

[10] Kikuo Maekawa. Corpus of Spontaneous Japanese: Its design and evaluation. In *Proc. of SSPR2003*, pages 7–12, 2003.

[11] Hiromitsu Nishizaki and Seiichi Nakagawa. Japanese spoken document retrieval considering OOV keywords using LVCSR system with OOV detection processing. In *Proc. of HLT2002*, pages 144–151, 2002.

[12] N. Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man and Cybernetics*, SMC-9(1):62–66, 1979.

[13] A. Park, Timoty J. Hazen, and James Glass. Automatic processing of audio lectures for information retrieval: Vocabulary selection and language modeling. In *Proc. of ICASSP2005*, pages 497–500, 2005.

[14] S. Togashi, M. Yamaguchi, and S. Nakagawa. Summarization of spoken lectures based on linguistic surface and prosodic information. In *Proc. of the IEEE/ACM Workshop on Spoken Language Technology(SLT)*, pages 34–37, 2006.

[15] Isabel Trancoso, Ricardo Nunes, Luís Neves, Céu Vianan, Helena Moniz, Diamonino Caseiro, and Ana Isabel Mata. Recognition of Classroom Lectures in European Portuguese. In *Proc. of Interspeech2006-ICSLP*, pages 281–284, 2006.

[16] J. Ueberla. Analysing a simple language model - some general conclusion for language models for speech recognition. *Computer Speech and Language*, 2(2):153–176, 1994.