# CENSREC-3: Data Collection for In-Car Speech Recognition and Its Common Evaluation Framework

Masakiyo Fujimoto[1], Satoshi Nakamura[1], Kazuya Takeda[2], Shingo Kuroiwa[3], Takeshi Yamada[4], Norihide Kitaoka[5], Kazumasa Yamamoto[6], Mitsunori Mizumachi[7], Takanobu Nishiura[8], Akira Sasou[9], Chiyomi Miyajima[2], and Toshiki Endo[1]

[1]ATR Spoken Language Translation Research Laboratories [2]Nagoya University
[3]University of Tokushima [4]University of Tsukuba [5]Toyohashi University of Technology
[6]Shinshu University [7]Kyushu Institute of Technology [8]Ritsumeikan University
[9]National Institute of Advanced Industrial Science and Technology

## Abstract

*This paper introduces a common database, an evaluation framework, and its baseline recognition results for in-car speech recognition, CENSREC-3, as an outcome of IPSJ-SIG SLP Noisy Speech Recognition Evaluation Working Group. CENSREC-3 which is a sequel of AURORA-2J is designed as the evaluation framework of isolated word recognition in real driving car environments. Speech data was collected using 2 microphones, a close-talking microphone and a hands-free microphone, under carefully controlled 16 different driving conditions, i.e., combinations of 3 car speeds and 5 car conditions. CENSREC-3 provides 6 evaluation environments which are designed using speech data collected in these car conditions.*

## 1 Introduction

The recent progress of speech recognition technology has been brought about by the advent of statistical modeling and large-scale corpora. Furthermore, it is also known that progress has been accelerated by the U.S. DARPA projects initiated in the late '80s in terms of project participants competitively developing speech recognition systems on the same task, using the same training and test corpus.

However, current speech recognition performance must still be improved if the system is to be exposed to noisy environments, where speech recognition applications might be used in practice. Therefore, robustness to acoustic noise is an emerging and crucial factor to be solved for speech recognition systems.

With regard to the noise robustness problem, there have been two evaluation projects, SPINE1, 2 [2] and AURORA [6]-[12]. The SPINE (SPeech recognition In Noisy Environment) project was organized by U.S. DARPA, with SPINE1 in 2000 and SPINE2 in 2001. The task included spontaneous English dialog between an operator and a soldier in a noisy field to evaluate spontaneous continuous speech recognition in noisy environments. The results of the project brought many improvements to continuous noisy speech recognition, though the task seems quite special and a little difficult to handle.

On the other hand, the European Telecommunications Standards Institute (ETSI) AURORA group initiated a special session in the EUROSPEECH conference. They are actively working to develop standard technologies under ETSI for distributed speech recognition [3]. In parallel with their standardization activities, they have distributed to academic researchers a noisy connected speech corpus based on TI-digits [4] with baseline HTK (HMM Took Kit) [5] scripts for further noisy speech recognition research. To date, AURORA2 [6]: a connected digit corpus with additive noise, AURORA3 [7]-[10]: an in-car noisy digit and word corpus, and AURORA4 [11, 12]: a large vocabulary continuous speech recognition with additive noise (noisy Wall Street Jounal, vocabulary size: 5,000) have been distributed with HTK scripts, which can be used to obtain baseline performance and relative improvements over the baseline results [13].

The authors voluntarily organized a special working group in October 2001 under the auspices of the Information Processing Society of Japan in order to assess speech recognition technology in noisy environments. The focus of the working group included the planning of comprehensive fundamental assessments of noisy speech recognition, standardized corpus collection, evaluation strategy developments, and distribution of standardized processing modules.

As an outcome of working group, we have already been produced the Japanese AURORA-2: AURORA-2J [14], which is translated English digits into Japanese.

This paper introduces a common database, an evaluation framework, and its baseline recognition results for in-car speech recognition, CENSREC-3 (Corpus and Environments for Noisy Speech RECognition), as a sequel of AURORA-2J [1]. The CENSREC-3 is designed as the evaluation framework of isolated word recognition in real driving car environments. Speech data was collected using 2 microphones, a close-talking microphone and a hands-free microphone, under carefully controlled 16 different driving conditions, i.e., combinations of 3 car speeds and 5 car conditions. The CENSREC-3 provides 6 evaluation environments which are designed using speech data collected in these car conditions. Finally, this paper shows the evaluation results of CENSREC-3 by using ETSI standard DSR front-end ES 202 050 [15], i.e., Advanced front-end.

## 2 Data recording

The CENSREC-3 database is composed of part of the database collected by the Center for Integrated Acoustic Information Research (CIAIR) [16].

### 2.1 Vocabulary

The speech recognition task of the CENSREC-3 database is isolated word recognition in real driving car environments. Table 1 shows a list of 50 words recorded for testing data. The acoustic models are trained by phonetically balanced sentences collected in real driving car environments.

**Table 1. A list of 50 recorded words.**

| | |
|---|---|
| digital_locker | ninsho_kaishi |
| 2001/1/1 | yamada_tarou |
| kensaku_shuryo | ansho_bango |
| 0123 | 4567 |
| 8901 | 2345 |
| 6789 | contents |
| eiga | Hitsuji_tachino_chinmoku |
| Sound_of_music | game |
| Pack_man | ongaku |
| jpop | konsyu_no_top10 |
| genre_betsu_kensaku | pops |
| rock | Beatles |
| senkyoku | Yesterday |
| Let_it_be | haishin_kaishi |
| ferry_annai | jikoku_hyo |
| dai2bin_wo_yoyaku | net_news |
| topics | onsei_yomiage |
| tenki_yohou | koutsu_jouhou |
| Kanagawa_ken | Yokohama_shi |
| Naka_ku | Toukyou_to |
| Setagaya_ku | Syuto_kousoku |
| Touhoku_jidoushadou | Seven_eleven |
| Uniqlo | Star_bucks |
| hotel_ichiran | Pacific_hotel |
| yoyaku_hyo | service_syuryo |

### 2.2 Speech data recording

In-car speech data was collected in a specially equipped vehicle. Six microphones were mounted to the vehicle as shown in Figure 1. Microphone no. 1 was a close-talking headset microphone, microphones no. 3 and 4 were attached to the dashboard, and microphones no. 5, 6, and 7 were fixed to the ceiling of the vehicle. The speech data recorded with the close-talking (CT) microphone (no.1: SONY ECM77B mounted on SENNHEISER HMD410) and the hands-free (HF) microphone attached to the ceiling of the driver's seat (no.6: SONY ECM77B) are used for CENSREC-3 [16].

The recording conditions for the evaluation data are shown in Table 2. Speech data was recorded under 16 environmental conditions using combinations of three kinds of vehicle speeds (idling, low-speed driving on a city street, and high-speed driving on an expressway) and six kinds of in-car environments (normal, with hazard flasher on, with air-conditioner on (fan low/high), with audio CD player on, and with windows open). A total of 14,216 utterances spoken by 18 speakers (8 males and 10 females) were recorded with each microphone.

**Table 2. Recording environments for testing data.**

| Car speed | In-car conditions |
|---|---|
| Idling (quiet) | Normal, Hazard on, Fan (low), Fan (high), Audio on, Window open |
| Low speed | Normal, Fan (low), Fan (high), Audio on, Window open |
| High speed | Normal, Fan (low), Fan (high), Audio on, Window open |

For training, driver's speech of phonetically-balanced sentences was recorded under two conditions: while idling and driving on a city street with a normal in-car environment [16]. A total of 14,050 utterances spoken by 293 drivers (202 males and 91 females) were recorded with each microphone. The drivers uttered the sentences by reading the written texts while idling. In the case of recording while driving, the sentences were divided into some

[1] AURORA-2J is regarded as a part of the CENSREC series and has been given an alternative name, CENSREC-1

**Table 3. Training data for each evaluation condition.**

| Evaluation condition | Condition 1 | | Condition 2 | | Condition 3 | | Condition 4 | | Condition 5 | | Condition 6 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Microphone | CT | HF | CT | HF | CT | HF | CT | HF | CT | HF | CT | HF |
| Idling (quiet) | ○ | ○ | ○ | — | — | ○ | — | ○ | ○ | — | ○ | — |
| Low speed | ○ | ○ | ○ | — | — | ○ | — | — | ○ | — | — | — |

**Table 4. Testing data for each evaluation condition.**

| Evaluation condition | Condition 1 | | Condition 2 | | Condition 3 | | Condition 4 | | Condition 5 | | Condition 6 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Microphone | CT | HF | CT | HF | CT | HF | CT | HF | CT | HF | CT | HF |
| Idling (quiet) | ○ | ○ | ○ | — | — | ○ | — | — | — | — | — | — |
| Low speed | ○ | ○ | ○ | — | — | ○ | — | ○ | — | ○ | — | ○ |
| High speed | ○ | ○ | ○ | — | — | ○ | — | ○ | — | ○ | — | ○ |

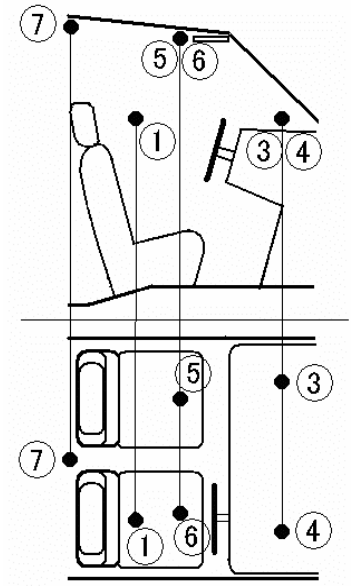


**Figure 1. Microphone positions for data collection: Side view (top) and top view (bottom).**

short segments to be easily memorized by the drivers. The drivers uttered each segment of the sentences after listening to the recorded instruction speech played via a headphone. Speech data of the segments were saved in separate files.

The speech signals for training and evaluation were both sampled at 16 kHz, quantized into 16 bit integers, and saved in the little-endian format.

## 3 Design of the evaluational framework

CENSREC-3 provides six evaluation environments for speech recognition using the speech data collected in various in-car conditions as described in the previous section [2]. Each evaluation framework consists of the conditions marked by a circle (○) in Tables 3 and 4. For each of conditions 1, 2, and 3, data collected by using the same microphones in the same recording environment were prepared both for training and testing. These conditions correspond to the "Well-matched condition" of the AURORA3 framework [7]-[10]. Condition 4 corresponds to the "Moderate-mismatched condition" of the AURORA3 framework, of which training and testing data were recorded under different conditions, that is, training and testing data were collected while idling and driving, by using the same microphones. Both condition 5 and condition 6 correspond to the "High-mismatched condition" of the AURORA3 framework, of which data collected by using different microphones under different recording conditions are used each for training and testing. Tables 5 and 6 show the amount of the data for training and testing in each condition.

## 4 Baseline performance

### 4.1 Baseline scripts for evaluation

The baseline scripts were designed to facilitate HMM training and evaluation by HTK [5]. The evaluation framework was designed as follows:

[2]Note that a license fee is required ONLY FOR part of the training data, which were collected by using a HANDS-FREE MICROPHONE. You should pay the license fee if you wish to use a part of the charged data collected by using distant-talking microphones, although the CENSREC-3 DVD disk includes both free and charged speech data,

**Table 5. The amount of training data for each evaluation condition.**

| Car speed | Microphone | Condition 1 | Condition 2 | Condition 3 | Condition 4 | Condition 5 | Condition 6 |
|---|---|---|---|---|---|---|---|
| Idling (quiet) | CT | 3,608 | 3,608 | — | — | 3,608 | 3,608 |
| | HF | 3,608 | — | 3,608 | 3,608 | — | — |
| | Total | **7,216** | **3,608** | **3,608** | **3,608** | **3,608** | **3,608** |
| Low speed | CT | 10,442 | 10,442 | — | — | 10,442 | — |
| | HF | 10,442 | — | 10,442 | — | — | — |
| | Total | **20,884** | **10,442** | **10,442** | — | **10,442** | — |
| Total | | **28,100** | **14,050** | **14,050** | **3,608** | **14,050** | **3,608** |

**Table 6. The amount of testing data for each evaluation condition.**

| Car speed | Microphone | In-car condition | Condition 1 | Condition 2 | Condition 3 | Condition 4 | Condition 5 | Condition 6 |
|---|---|---|---|---|---|---|---|---|
| Idling (quiet) | CT | Normal | 898 | 898 | — | — | — | — |
| | | Hazard on | 900 | 900 | — | — | — | — |
| | | Fan (low) | 887 | 887 | — | — | — | — |
| | | Fan (high) | 900 | 900 | — | — | — | — |
| | | Audio on | 896 | 896 | — | — | — | — |
| | | Window open | 899 | 899 | — | — | — | — |
| | | Total | 5,380 | 5,380 | — | — | — | — |
| | HF | Normal | 898 | — | 898 | — | — | — |
| | | Hazard on | 900 | — | 900 | — | — | — |
| | | Fan (low) | 887 | — | 887 | — | — | — |
| | | Fan (high) | 900 | — | 900 | — | — | — |
| | | Audio on | 896 | — | 896 | — | — | — |
| | | Window open | 899 | — | 899 | — | — | — |
| | | Total | 5,380 | — | 5,380 | — | — | — |
| | Total | | **10,760** | **5,380** | **5,380** | — | — | — |
| Low speed | CT | Normal | 848 | 848 | — | — | — | — |
| | | Fan (low) | 850 | 850 | — | — | — | — |
| | | Fan (high) | 895 | 895 | — | — | — | — |
| | | Audio on | 849 | 849 | — | — | — | — |
| | | Window open | 897 | 897 | — | — | — | — |
| | | Total | 4,339 | 4,339 | — | — | — | — |
| | HF | Normal | 848 | — | 848 | 848 | 848 | 848 |
| | | Fan (low) | 850 | — | 850 | 850 | 850 | 850 |
| | | Fan (high) | 895 | — | 895 | 895 | 895 | 895 |
| | | Audio on | 849 | — | 849 | 849 | 849 | 849 |
| | | Window open | 897 | — | 897 | 897 | 897 | 897 |
| | | Total | 4,339 | — | 4,339 | 4,339 | 4,339 | 4,339 |
| | Total | | **8,678** | **4,339** | **4,339** | **4,339** | **4,339** | **4,339** |
| High speed | CT | Normal | 900 | 900 | — | — | — | — |
| | | Fan (low) | 900 | 900 | — | — | — | — |
| | | Fan (high) | 900 | 900 | — | — | — | — |
| | | Audio on | 899 | 899 | — | — | — | — |
| | | Window open | 898 | 898 | — | — | — | — |
| | | Total | 4,497 | 4,497 | — | — | — | — |
| | HF | Normal | 900 | — | 900 | 900 | 900 | 900 |
| | | Fan (low) | 900 | — | 900 | 900 | 900 | 900 |
| | | Fan (high) | 900 | — | 900 | 900 | 900 | 900 |
| | | Audio on | 899 | — | 899 | 899 | 899 | 899 |
| | | Window open | 898 | — | 898 | 898 | 898 | 898 |
| | | Total | 4,497 | — | 4,497 | 4,497 | 4,497 | 4,497 |
| | Total | | **8,994** | **4,497** | **4,497** | **4,497** | **4,497** | **4,497** |
| Total | | | **28,432** | **14,216** | **14,216** | **8,836** | **8,836** | **8,836** |

- All scripts are written in Perl, and work with Perl version 5 and later.

- The CENSREC-3 database provides parallel processing by multiple computers to reduce the processing time. The parallel processing is easily available by simply adding the remote host names to the configuration file of the baseline scripts.

- The speech recognition is carried out using phoneme HMMs. In the recognition, a standard pronunciation dictionary and recognition grammar described by the EBNF syntax notation as shown in Figure 2 are defined.

- The acoustic models consist of triphone HMMs that have five states with three distributions. Each distribution is represented with 32 mixture Gaussians. The total number of states that have the distributions are 2,000.

- In the case of a word with connected vowels that can be pronounced by a long vowel, pronunciation rules for both the connected vowels and the long vowel are registered in the pronunciation dictionary. For example, in the case of the Japanese word "Ninshou", two pronunciation rules, "n i N sh **o u**" and "n i N sh **o:**", are registered.

- The feature vector consisted of 12 MFCCs and log-energy with their corresponding delta and acceleration coefficients. Analysis conditions were pre-emphasis $1-0.97z^{-1}$, hamming window, 20-msec frame length, and 10-msec frame shift. In the baseline performance, cepstral mean subtraction was not applied to the feature vectors.

- In the Mel-filter bank analysis, a cut-off was applied to frequency components lower than 250 Hz.

```
$words = digital_locker |
         ninsho_kaishi |
         ... |
         service_syuryo;

( [silB] $words [silE] )
```

**Figure 2. Grammar written in EBNF.**

### 4.2 Baseline recognition results and performance comparison

Table 7 shows the details of baseline recognition results for each car environment for evaluation conditions 1 to 6 [3].

We will also distribute a Microsoft Excel spreadsheet to simplify the recognition performance comparison. All of the baseline results and the averaged recognition result are shown in the top of Table 8. The data entry for your results (word accuracy) should be made in the middle part of Table 8, then the relative improvement against the baseline result is automatically given in the bottom part.

Table 8 also shows the evaluation results of ETSI ES 202 050 front-end. In the table, we can see that the results by ETSI ES 202 050 front-end are considerably higher than that of the baseline performance.

## 5 Evaluation categories

Evaluation categories are designed for CENSREC-3, which show how much the user's method modified the baseline back-end scripts from the viewpoint of changes in the training method of HMMs, model topology, feature parameters, and so on. Users are requested to declare the category to which they belong from the following categories according to the degree of modification to the back-end scripts from the original baseline. No changes to the back-end scripts, i.e., changes to only front-end processing, can be included in category 0. Recognition results can be fairly compared with other methods only within the same category. In addition, the following categories are from AURORA-2J with some changes.

**Category 0.** No changes to the back-end scripts.

**Category 1.** If the HMM topology is the same as the baseline scripts, any training process will be allowed. Discriminative training can be introduced in this category. The computational cost in the recognition phase should be the same as it was. Other experimental conditions are the same as in the back-end scripts.

**Category 2.** If the HMM topology is the same, adaptation processes can be introduced using some testing data. Speaker or environment adaptation, and PMC with one state noise model can be allowed in this category. An increase in the computational cost will be caused only

[3] There may be cases where the parameters of acoustic models change slightly according to the number of computers and the operating system used for experiments. This often affects the recognition results (its fluctuation is approximately $\pm 1\%$). The experiments for obtaining the baseline results were performed by using four computers with Red Hat Linux release 7.2. This phenomenon has repeatability. Hence, when you carry out the baseline evaluation with four computers, you can obtain the same results as shown in Table 7.

**Table 7. Details of CENSREC-3 baseline evaluation results (%).**

| Car speed | Microphone | In-car condition | Condition 1 | Condition 2 | Condition 3 | Condition 4 | Condition 5 | Condition 6 |
|---|---|---|---|---|---|---|---|---|
| Idling (quiet) | CT | Normal | 99.89 | 100.00 | — | — | — | — |
| | | Hazard on | 99.33 | 99.89 | — | — | — | — |
| | | Fan (low) | 99.55 | 100.00 | — | — | — | — |
| | | Fan (high) | 97.78 | 99.44 | — | — | — | — |
| | | Audio on | 98.77 | 99.67 | — | — | — | — |
| | | Window open | 99.11 | 99.33 | — | — | — | — |
| | | Overall | 99.07 | 99.72 | — | — | — | — |
| | HF | Normal | 99.44 | — | 99.78 | — | — | — |
| | | Hazard on | 98.78 | — | 98.89 | — | — | — |
| | | Fan (low) | 90.19 | — | 94.02 | — | — | — |
| | | Fan (high) | 53.56 | — | 53.44 | — | — | — |
| | | Audio on | 81.47 | — | 81.36 | — | — | — |
| | | Window open | 89.66 | — | 89.88 | — | — | — |
| | | Overall | 85.50 | — | 86.21 | — | — | — |
| | Overall | | **92.29** | **99.72** | **86.21** | — | — | — |
| Low speed | CT | Normal | 100.00 | 100.00 | — | — | — | — |
| | | Fan (low) | 100.00 | 100.00 | — | — | — | — |
| | | Fan (high) | 97.99 | 98.77 | — | — | — | — |
| | | Audio on | 98.82 | 99.41 | — | — | — | — |
| | | Window open | 99.11 | 98.55 | — | — | — | — |
| | | Overall | 99.17 | 99.33 | — | — | — | — |
| | HF | Normal | 98.00 | — | 99.17 | 88.21 | 56.60 | 45.99 |
| | | Fan (low) | 90.82 | — | 94.12 | 77.41 | 54.35 | 35.18 |
| | | Fan (high) | 62.57 | — | 60.11 | 41.79 | 43.46 | 28.83 |
| | | Audio on | 79.27 | — | 78.56 | 65.02 | 47.47 | 37.57 |
| | | Window open | 64.66 | — | 65.33 | 45.60 | 23.97 | 15.27 |
| | | Overall | 78.73 | — | 79.10 | 63.17 | 44.92 | 32.33 |
| | Overall | | **88.95** | **99.33** | **79.10** | **63.17** | **44.92** | **32.33** |
| High speed | CT | Normal | 99.89 | 99.89 | — | — | — | — |
| | | Fan (low) | 99.67 | 99.89 | — | — | — | — |
| | | Fan (high) | 97.67 | 99.22 | — | — | — | — |
| | | Audio on | 99.78 | 99.78 | — | — | — | — |
| | | Window open | 96.66 | 95.21 | — | — | — | — |
| | | Overall | 98.53 | 98.80 | — | — | — | — |
| | HF | Normal | 92.33 | — | 95.56 | 64.78 | 29.67 | 21.78 |
| | | Fan (low) | 85.11 | — | 89.44 | 48.22 | 30.67 | 19.89 |
| | | Fan (high) | 59.67 | — | 55.22 | 37.33 | 40.78 | 22.44 |
| | | Audio on | 78.31 | — | 79.20 | 49.72 | 30.03 | 23.92 |
| | | Window open | 24.83 | — | 21.83 | 15.37 | 7.80 | 6.46 |
| | | Overall | 68.07 | — | 68.27 | 43.10 | 27.80 | 18.90 |
| | Overall | | **83.30** | **98.80** | **68.27** | **43.10** | **27.80** | **18.90** |
| Overall | | | **88.43** | **99.31** | **78.36** | **52.95** | **36.20** | **25.50** |

by the adaptation process. Other experimental conditions are the same as in the back-end scripts.

**Category 3.** Changes in the standard HMM topology. A different number of mixtures and states can be allowed. However, the recognition unit should be the same as in the original back-end scripts ("triphone HMMs" in CENSREC-3). PMC with more than one state noise model can be included in this category. Other experimental conditions are the same as in the back-end scripts.

**Category 4.** Any process will be allowed as long as the decoder is the same as in the original back-end scripts (`HVite` in CENSREC-3). Changes of a model unit, syntax and lexicon for the decoder can be included in this category.

**Category 5.** Any process with any computational cost will be allowed.

**Category B.** The use of any training data not included in CENSREC-3 — not only speech data, but also environment noise data. Of course, the evaluation data is CENREC-3. This category essentially differs from categories 1 to 5.

**Table 8. CENSREC-3 spreadsheet and the evaluation results of ETSI ES 202 050 front-end.**

| CENSREC-3 Evaluation Results | | | | | | |
|---|---|---|---|---|---|---|
| CENSREC-3 Baseline Results (%) | | | | | | |
| Condition 1 | Condition 2 | Condition 3 | Condition 4 | Condition 5 | Condition 6 | Average |
| 88.43 | 99.31 | 78.36 | 52.95 | 36.20 | 25.50 | 63.46 |
| CENSREC-3 Word Accuracy (%) | | | | | | |
| Condition 1 | Condition 2 | Condition 3 | Condition 4 | Condition 5 | Condition 6 | Average |
| 95.48 | 99.62 | 91.95 | 86.63 | 83.70 | 73.85 | 88.54 |
| CENSREC-3 Relative Improvement | | | | | | |
| Condition 1 | Condition 2 | Condition 3 | Condition 4 | Condition 5 | Condition 6 | Average |
| 60.93% | 44.93% | 62.80% | 71.58% | 74.45% | 64.90% | 68.63% |

## 6 Conclusions

In this paper, we introduced CENSREC-3, an evaluation framework for Japanese in-car speech recognition and showed evaluation results of ETSI ES 202 050 front-end.

In the near future, we will develop the series of frameworks for noisy speech recognition, CENSREC-1.5 (AURORA-2.5J): a subset of AURORA-2J with the Lombard effect speech and CENSREC-2 (AURORA-3J): continuous digits speech database collected in real driving car environments.

We also plan to design and distribute the evaluation frameworks of noisy speech recognition gradually made difficult, i.e., non-stationary noise environments, reverberant environments, large vocabulary continuous speech recognition task, and so on. Furthermore, we plan to develop and distribute a noise database for noisy speech recognition, evaluation measures instead of the word accuracy, and a tool kit of conventionally used noise compensation methods.

We will give the latest information about CENSREC in the following Web site.

**CENSREC Web site:**
http://sp.shinshu-u.ac.jp/CENSREC/

## Acknowledgements

The research reported here was supported in part by a contract with the National Institute of Information and Communications Technology entitled “A study of speech dialogue translation technology based on a large corpus”.

## References

[1] DARPA project Web site, http://www.nist.gov/speech/publications/

[2] SPINE Web site, http://elazar.itd.nrl.navy.mil/spine/

[3] ETSI Web site, http://www.etsi.org/

[4] R. G. Leonard, “A database for speaker independent digit recognition”, Proc. ICASSP’84, vol. 3, pp. 328-331, 1984.

[5] HTK Web site, http://htk.eng.cam.ac.uk/

[6] H.G.Hirsch and D.Pearce, “The AURORA Experimental Framework for the Performance Evaluations of Speech Recognition Systems under Noisy Condition”, Proc. ISCA ITRW ASR2000, pp. 18-20, Paris, France, Sep. 2000.

[7] AU/378/01, “Danish SpeechDat-Car Digits Database for ETSI STQ-Aurora Advanced DSR”, Aalborg University, Jan. 2001.

[8] AU/225/00, “Baseline Results for subset of SpeechDat-Car Finnish Database for ETSI STQ WI008 Advanced Front-end Evaluation”, Nokia, Jan. 2000

[9] AU/273/00, “Description and Baseline Results for the Subset of the Speechdat-Car German Database used for ETSI STQ Aurora WI008 Advanced DSR Front-end Evaluation”, Texas Instruments, Dec. 2001.

[10] AU/271/00, “Spanish SDC-Aurora Database for ETSI STQ Aurora WI008 Advanced DSR Front-End Evaluation: Description and Baseline Results”, UPC, Nov. 2000.

[11] AU/337/01, “Experimental Framework for the Performance Evaluation of Speech Recognition Front-Ends on a Large Vocabulary Task: Version 1.0”, Ericsson, June 2001.

[12] AU/345/01, “Large Vocabulary Evaluation of Front-ends: Baseline Recognition System Description, Final Report”, Mississippi State University, Jan. 2002.

[13] D. Pearce, “Developing the ETSI AURORA advanced distributed speech recognition front-end & What next”, Proc. EUROSPEECH2001, Sep. 2001.

[14] S.Nakamura, K.Takeda, K.Yamamoto, T.Yamada, S.Kuroiwa, N.Kitaoka, T.Nishiura, A.Sasou, M.Mizumachi, C.Miyajima, M.Fujimoto, and T.Endo, “AURORA-2J: An Evaluation Framework for Japanese Noisy Speech Recognition”, IEICE Transactions on Information and Systems, Vol.E88-D, No.3, Mar. 2005. (to appear)

[15] ETSI ES 202 050 V1.1.1, “Distributed speech recognition; advanced front-end feature extraction algorithm; compression algorithms”, 2002.

[16] K.Takeda, H.Fujimura, K.Itou, N.Kawaguchi, S.Matsubara, and F.Itakura, “Construction and Evaluation a Large In-Car Speech Corpus”, IEICE Transactions on Information and Systems, Vol. E88-D, No. 3, Mar. 2005. (to appear)