

The Present Status of Speech Database in Japan: Development, Management, and Application to Speech Research

Hisao Kuwabara¹⁾, Shuich Itahashi²⁾, Mikio Yamamoto²⁾, Toshiyuki Takezawa³⁾,
Satoshi Nakamura³⁾, Kazuya Takeda⁴⁾

1) Department of Electronics and Information Science, Teikyo University of Science & Technology
Uenohara, Kitatsuru-gun, Yamanashi 409-0193, Japan,

kuwabara@ntu.ac.jp

2) Institute of Information Sciences and Electronics, Tsukuba University
Tennodai, Tsukuba, Ibaraki 305-8573, Japan,

itahashi@milab.is.tsukuba.ac.jp, myama@is.tsukuba.ac.jp

3) ATR Spoken Language Translation Research Laboratories
Hikaridai, Soraku-gun, Kyoto 619-0288, Japan,

{takezawa, nakamura}@slt.atr.co.jp

4) Center for Integrated Acoustic Information Research, Nagoya University
Furomachi, Chigusa-ku, Nagoya 464-8603, Japan,

takeda@nuee.nagoya-u.ac.jp

Abstract

The present status of Japanese speech database has been described. The database project in Japan started in early 1980s. The first one was a committee of Japan Electronic Industry Development Association, abbreviated as JEIDA, which aimed at creating a speech database that can commonly evaluate performance of the then existing speech input/output machines and systems. Several database projects have been undertaken since then including the one initiated by the Advanced Telecommunication Research Institute (ATR) and now it has come to the point where an enormous amount of spontaneous speech data is available. A survey has been conducted recently about the usage of the presently existing speech databases among industry and university institutions in Japan where speech research is now actively going on. It has been revealed that the ATR's continuous speech database is the most frequently used followed by the equivalent version of the Acoustical Society of Japan.

1. Introduction

As the activities of speech research not only in the field of speech technologies but in many academic areas including linguistics and physiology increase, a demand for large-scale speech databases for common use and many purposes grows. Before common speech databases were available, people were making their own small databases for their own use in almost every institution without distributing outside. As the computer facilities go up, especially computer's processing speed and memory size, it has become possible to handle a large amount of speech data. Recent speech recognition and synthesis technologies, such as HMM, N-gram language models and corpus-base synthesis, require an enormous amount of speech data for both training and assessment. It was from this atmosphere that collection and building a large-scale speech database began in Japan years ago and now it comes to a point where a great deal of spontaneous speech data are available.

Speech database projects in the so-called industrially advanced countries have been stimulated to grow more and more active from the late 1980s by a series of database workshops, i.e. in The Netherlands in 1989 (ESCA Workshop, 1989), in Japan in 1990 (ICSLP Workshop, 1990) and in Italy in 1991 (ESCA Workshop, 1991). In the U.S., the most advanced country in the database project, has already started LDC in the early 1990s based on the accumulation of databases collected through ARPA

and DARPA projects. In Europe, they have created ELRA in 1995 as the non-profit organization. It collects, markets and distributes language resources along with the dissemination of general information in the field of language engineering. In Japan, there is a new project called GSK, created in 1999, which aims at accumulating and controlling language resources. However, unlike LDC and ELRA, the GSK does not include speech corpora yet while the former two are the integrated bodies for both text and speech data. The present situation in organizing the database projects in Japan is still a little behind other advanced countries.

Speech database project in Japan started in early 1980s initiated by JEIDA (Japan Electronic Industry Development Association). It was a time when several high-tech firms began delivering speech I/O machines to the market claiming the best performance by themselves. There was a need to standardize the evaluation of these I/O machines using common speech database (Itahashi, 1999). Soon after finishing this initial project in 1985, ATR (Advanced Telecommunication Research Institute, International) database construction began from the very beginning of its foundation (Kuwabara, 1989). ASJ (Acoustic Society of Japan) database project has been created in mid 1990s by re-forming the previous JEIDA project and adding more members in the committee and still continues its activity. This paper describes the present status of Japanese speech database with a result of recently conducted questionnaire to industry/university institutions

about the usage of speech database.

2. A Brief History of Database Project

As mentioned above, the database project in Japan began through the formation of JEIDA organization in 1982 headed by Prof. S. Itahashi of Tsukuba University. It was a time when Japan needed a common speech database that would be used to uniformly evaluate the then existing speech recognition machines and systems. The corpus constructed was relatively small consisting of about 300 isolated words with 150 speakers, 75 for male and female each (Itahashi, 1985). Speakers' ages were carefully regulated so that they would constitute approximately the same age-ratio as the whole Japanese population at that time. This database was used in many institutions to evaluate their own products and the project finished its first mission in 1985 to succeed to another one.

The JEDA committee was re-formed to create a new organization attached to ASJ (Acoustical Society of Japan) in 1990 adding more committee members and headed again by Prof. S. Itahashi. Under the new committee, several projects have been undertaken to create speech corpora placed more weights on basic speech input/output technologies and research than before. Under this committee, several speech corpora were created and distributed which will be mentioned later in this paper. The committee, the first author of this paper succeeded chairman three years ago, still continues and is responsible for distributing/managing the existing databases and for future planning.

ATR (Advanced Telecommunication Research Institute International) started its database project at the same time of its foundation in 1986 and the first author of this paper was responsible for it. It was time when new speech recognition technologies such as HMM, Neural Network and Language Models emerged. These technologies require a large amount of speech data both for training and for evaluation. Initially, it intended to collect clean read speech with professional announcers and narrators. It consists of 1) word speech database and 2) continuous speech database, and the most important aspect of this database is that it provides detailed labels in five layers (Takeda, 1988). ATR's database project still continues today under the name ATR-SLT (ATR Spoken Language Translation Research Laboratories) database.

A four-year database project was running from 1989 to 1992 under the priority area on "Prosodic Features of Spoken Japanese" sponsored by the then Ministry of Education (Kuwabara, 1996). The purpose of this project was to create a speech corpus by collecting regional and dialectic spoken Japanese. It aims at providing a useful database mainly for linguists, phoneticians, dialectologists, speech therapists and language educators. Those who participated in this project numbered more than 300 but most of them from outside speech engineering.

Recently, a few government-sponsored projects related to speech research/database are now underway in Japan. First, a five-year project called "Spontaneous Speech Engineering," started in 1999, is under way (Maekawa, 2000). This is a joint project connecting The National Language Research Institute, Tokyo Institute of Technology, and Communications Research Laboratory headed by Prof. S. Furui of Tokyo Institute of Technology. In this project, they are trying to build a corpus of spontaneous

Japanese called CSJ (Corpus of Spontaneous Japanese). However, what they are collecting is not "spontaneous" speech in the exact sense. They are collecting so-called "monologue" speech by recording academic presentation's speech and simulated public speech. It aims at investigating linguistic, para-linguistic structure of spoken Japanese. It has collected so far more than 680 hours (7.8 million words) of monologue speech data and approximately 520 hours of them were transcribed into written text.

Second, another five-year project headed by Prof. F. Itakura of Nagoya University started in the same 1999. This is "Integrated Acoustic Information Research" project. It aims at integrating different aspects of human-sound relation by investigating not only speech sounds but also other acoustic signals. Reflecting Japan's aging society, a new speech corpus consisting of 400 elderly people's speech has been completed recently under this project.

Third, a four-year project on "Realization of Advanced Spoken Language Information Processing for Prosodic Features," headed by Prof. K. Hirose of Tokyo University, has now entered to the third year. This project, to which some of the authors of this paper join, aims at developing and researching advanced speech input/output technologies based mainly on the prosodic features of spoken Japanese. In this project, a spontaneous-speech corpus has now being constructed mainly for prosody research and text-to-speech synthesis (Kitazawa, 2001).

3. The Present Speech Corpora

As mentioned above, there are several speech corpora in Japan created in the course of progressing individual database projects. Followings are the existing speech databases in Japan including ones that are in the process of building and not yet completed.

3.1. JEIDA Isolated-Word Corpus

As mentioned above, the JEDA speech corpus was designed to commonly evaluate speech recognizers for commercial use at that time. Table 1 represents the content of the corpus. In the table, the "foreign syllable" stands for the pronunciations that are not in the original

Type of Words	Number
isolated syllable	101
foreign syllable	9
single digit	15
continuous digit	35
function word	63
city/place name	100
total	323

Table 1: Content of JEIDA isolated-word corpus

Japanese language but come from foreign words. The "single digit" stands for ten digit plus another pronunciations for "0, 4, 7, 9." The continuous digits are strings of four digits. The number of speakers are 150 with 75 male/female each.

3.2. ASJ Continuous Speech Corpus

The kinds of speech corpora, 1) "Continuous Speech

Corpus for Research (CSCR),” and 2) “Japanese Newspaper Article Sentences (JNAS),” were created so far in the ASJ speech database project. They are both read speech and the former is a mixture of sentence-read speech and simulated dialogue speech. Table 2 shows the content of ASJ speech corpus.

3.2.1. ASJ-CSCR

Table 2 is the content of CSCR database. Phonetically balanced sentences used here are the same as those of

Type of Sentences	Number	Speakers
phonetically balanced	503	64
guidance task	1027	36
simulated dialog	(37)	37

Table 2: Content of ASJ continuous speech corpus.

ATR database but different speakers. The “simulated dialog,” refers to the number of dialogues, not the number of sentences. Male and female speakers are almost equal number for all sentence types. These data are stored in 7 CD-ROMs. Over twenty institutions affiliated with industries and universities joined this project and collected speech data at their individual places.

3.2.2. SAJ-JNAS

JNAS corpus refers to “Japanese Newspaper Article Sentences.” This project was supported by Mainichi Shinbun, one of the leading daily newspapers in Japan, and the articles were offered by the Newspaper. Project members were the same as above mentioned CSCR project. The articles cover a variety of topics from politics to sports events. Approximately 15,500 sentences plus 50 ATR phonetically balanced sentences were prepared to read by 306 speakers with 153 male/female each. Table 3 shows the content. The data are stored in 16 CD-ROMs.

Type of Sentence	Number	Speaker
newspaper article	15,500	306
phonetically balanced	50	306

Table 3: Content of ASJ-JNAS speech corpus.

3.3. ATR Speech Corpus

ATR began its database project at the same time of its foundation in 1986. The first author of this paper participated in the project from the beginning. It was time when new speech methods like HMM, neural network, N-gram language models had emerged into the speech recognition technologies. These methods require a large amount of speech data for the training and also for the evaluation. A lot of discussions had been made about the database size, contents, speakers, and labeling. There are three kinds of database sets and data collection started with recordings of professional announcer’s voice in the noise-free environments. Data collection still continues today in the present ATR-SLT (ATR Spoken Language Translation Research Laboratories).

What is important in ATR speech database is that it gives data the following detailed labels in five layers;

- Layer 1: phonetic symbols
- Layer 2: acoustic events

- Layer 3: allophonic variants
- Layer 4: inseparable portions
- Layer 5: vowel centers

3.3.1. Isolated Word Database Set (Set A)

Words of this dataset were selected from a Japanese dictionary plus small number of phonetically balanced words, numerals, and others. Table 4 shows the content of

Words	Number	Speakers
important words from dictionary	5240	20
phonetically balanced words	216	20
alphabets and numerals	110	20
simulated dialog	(115)	20

Table 4: Content of ATR Isolated Word Database Set.

this dataset. The speakers are all professional announcers and the number is as of 1992 data. The “simulated dialog” is a task for travel information.

3.3.2. Continuous Speech Database Set (Set B)

This dataset is a read version of phonetically balanced 503 short Japanese sentences selected from newspapers and magazines (Iso, 1988). For the sake of simplicity, the sentences are grouped into ten sets, nine sets contain 50 sentences each plus one set 53, featuring with phonetically balanced within each set too. Speakers are 10 professional announcers with 5 male/female each.

3.3.3. Dataset for Large Number of Speakers (Set C)

This dataset was designed for use in the speaker independent speech recognition and in the speaker adaptation technologies. The content of this dataset is a mixture of

Item	Number	Speakers
words from Set A	750	250
sentences from Set B	150	250

Table 5: Content of dataset for large number of speakers

Set A and Set B; 750 words from Set A plus 150 sentences, also phonetically balanced, from Set B. Table 5 represents the content of this dataset. The number of speakers is 250 as of 1992 and still growing.

3.4. RWCP Speech Corpora

RWCP refers to as “Real World Computing Partnership” and the speech corpora are designed to meet developing speech technologies that can be used in our living environments, not in the laboratory environment. This started in 1992 as a ten-year project There are three datasets in the RWCP project, 1) RWCP sound scene database in real acoustic environments (Nakamura, 1999, 2000), 2) RWCP spoken dialogue dataset, 3) RWCP broadcasting news speech dataset, and 4) table-meeting speech dataset.

3.4.1. RWCP Sound Scene Database

Sound scene database was created for studies such as sound-source localization, sound retrieval, sound recog-

dition, and speech recognition in real acoustic environments. The hands-free recognition requires the following three major technologies to overcome fundamental problems.

- robustness to directional and omni-directional noises in the room.
- robustness to acoustic reflection and reverberation in the room.
- localization, tracing and recognition of a speaker among many sound sources including other speakers and noise.

In this project, they have collected three kinds of sounds and speech datasets. First, what they call “dry source sound,” a noise database was created to simulate the real office’s sound environment. In this dataset, various sounds that we encounter in the real living condition were recorded. Table 6 shows the “dry source sound.” Second, an impulse response database through microphone array with different positions in the room and different reverberation time was created. And third, a moving sound source database that the recording was made while the sound-source was moving from one position to another in the room.

3.4.2. RWCP Spoken Dialogue Database

This is a simulated dialogue database on two topics, about 1) car dealing and 2) traveling abroad. The dialogues were made in the question-answer style. In each topic, 24 dialogues were exchanged between the questioner and the respondent. The questioners are 24, twelve male/female each, and the respondents are 2 professionals in each field with a male and a female.

3.4.3. RWCP Broadcasting News Speech Database

This is a read-text version of broadcasting news articles by professional announcers. Six announcers participated in the readings and read 40 news-articles each from 190 listed articles. Each article contains approximately 490 Japanese alphabets.

3.4.4. RWCP Table-Meeting Speech Database

This is a speech database recorded in a meeting room when 4 or 5 people are sitting around a round-table and talking with each other wearing a small microphone each. The recording was made about 7 topics, approximately 30 minutes each, together with video pictures.

4. Speech Corpora under the Running Research Projects

There are four government-funded, though from different sections, speech research projects running now in Japan. They are 1) Spontaneous Speech Engineering: Corpus and Processing Technology, 2) Integrated Acoustic Information Research, 3) Realization of Advanced Spoken Language Information Processing from Prosodic Features, and 4) The Expressive Speech Processing Project (Campbell, 2001). In each project, creation of a speech database is planned. The followings are the brief introduction of some speech database schemes.

4.1. Corpus of Spontaneous Speech Engineering ? CSJ database ?

Type of Sound	Category	#samples
Collision Sound	wood	1187
	metal	1000
	plastic	550
	ceramic	800
Action Sound	article dropping	200
	gas jetting	200
	rubbing	500
	bursting/breaking	200
	clapping	829
Characteristic Sound	small metal articles	1072
	paper	400
	music instruments	1079
	electronic sounds	705
	mechanical sounds	1000

Table 6: RWCP source sounds for simulation of the real environment sound.

The CSJ refers to as “Corpus of Spontaneous Japanese,” and the project, started in 1999, has three main goals; 1) building a large-scale spontaneous speech corpus, 2) developing acoustic and linguistic modeling for spontaneous speech recognition, understanding and summarization, and 3) constructing a prototype system for spontaneous speech summarization (Furui, 2002). It has focused on the ‘monologue’ speech data rather than “dialogue” and collected so far a large amount of monologue data. There are two ways to collect the monologue data, 1) collection of academic presentation speech and 2) collection of simulated public speech.

Table 7 shows the contents of monologue data so far collected today (Maekawa, 2002). The third item in the

Type of Monologue	Amount (hours)
simulated public speech	325
academic presentation	310
dialogue/interview	34

Table 7: The content of CSJ collected so far.

table “dialogue/interview” was not involved in the original corpus plan but added later on as the data collection was unexpectedly ahead the schedule to some extent.

In this project, they are now giving data some linguistic and acoustic information, 1) transcriptions and 2) morphological annotations. Two kinds of transcriptions, one is orthographic and the other is phonetic, are being labeled in the data together with the information of utterance boundary, filler, dis-fluency, and noise. Unlike the read-speech, the spontaneous data, even though is monologue, are full of undesired utterances, such as fillers, noise, wrong or uncertain pronunciations, and so on. To cope with this, they have given 15 tags to indicate specific utterances. Morphological annotations include segmental and intonation labeling. All these tasks are underway in the National Language Research Institute and they try to finish the morphological annotations by the end of this project.

4.2. Corpus of the Integrated Acoustic Information Research

This project is headed by Prof. Itakura of Nagoya Uni-

versity and aims at integrating different aspects of human-sound relation on 1) the spatial sound localization by humans, 2) analysis and synthesis of sounds, 3) recognition of sounds including speech, 4) machine understanding of spoken language, and 5) human perception of sounds. This project plans to create large scale speech corpora and acoustic sound databases to be used in the project and also by the speech research community. So far, it has created a corpus of elderly people reflecting the recent trend towards the aging society.

In this project, they have created 7 databases for sound and speech. They are 1) database of room acoustics (CIAIR-RAM1999), 2) database of human head transfer functions (CIAIR-HRM1999), 3) database of driver's commands to in-vehicle navigators (CIAIR-HCC), 4) database of children's speech while playing video game (CIAIR-VCV), 5) database of simultaneous interpreters and original speakers speech (CIAIR-SI2000), 6) sounds of musical and acoustical instruments (CIAIR-AIS), and 7) database of elderly people (S-JNAS).

S-JNAS database consists of two datasets, 1) for acoustic modeling and 2) for recognition evaluation. The former one contains a total of 301 elderly peoples' speech for older than 60, and two persons are even older than 85. The latter one consists of 101 elder peoples' speech with again older than 60.

4.3. Corpus of Advanced Spoken Language Processing ? Japanese MULTEXT ?

In the project "Realization of Advanced Spoken Language Processing from Prosodic Features," they try to build a prosodic corpus (Kitazawa, 2001). The recordings were made based on the read-passages translated from the ELRA's MULTEXT prosodic corpus. The text was translated from five European languages and modified into Japanese. It comprises 40 different passages with 6523 morae including 1085 accent kernels.

Preliminary recordings have been made for 6 professional narrators and actors, 3 male/female each. For each speaker, recordings have been made in two ways; 1) read the text naturally, and 2) read it with an instructed emotional and paralinguistic attitude. Signals from EGG electrodes have also been recorded simultaneously. A total of 3 hours and 37 minutes speech have been recorded so far.

4.4. Corpus of Expressive Speech Processing ? JST/CREST database ?

This project aims at creating a very large spontaneous speech database, including emotional speech (Campbell, 2001). They try to record spontaneous samples, conversational speech and daily spoken interaction, of naturally-occurring emotions without recourse of acting. They try to avoid laboratory regulated speech samples. The project, funded by the Agency of Science and Technology, is now running jointly with 6 groups, 1) NAIST (Nara Institute of Science and Technology), 2) Kobe University, 3) Keio University, 4) Chiba University, 5) ICP Grenoble, and 6) ATR-ISD. The target is 1000 hours natural daily conversational speech data.

5. Database Utilization in Japan

Recently, a survey about the use of speech database has been conducted. By mailing questionnaire sheets to ap-

proximately 200 industry/university laboratories and collected 58 responses so far. The questions include:

- (1) Which database are you now using and for what purpose?
- (2) Which database did you use over the past 5 years and for what purpose?
- (3) Which database will you use in the near future and for what purpose?
- (4) Please give us information about the database you are now building or planning to build in the near future, if any.
- (5) What kind of database will be needed/necessary for the future speech research and technology development?

The respondents were allowed to answer the questions anonymously so that, especially for some people in industry, they would be able to respond more freely. The response data were analyzed with industry, government institution, university, and unknown (anonymous) groups separately.

Database	Use
JEIDA	12
RWCP	7
ATR	30
ASJ	22
TIMIT + TIDIGIT	6
CSJ	8
Self-made	12
Others	14

Table 8: Currently used database in Japan.

Database	Univ.	Ind.	Gov.	Unkn.
JEIDA	8	3	0	1
RWCP	5	0	1	1
ATR	14	8	3	5
ASJ	16	1	1	4
TIMIT + TIDIGIT	2	2	0	2
CSJ	6	2	0	0
Self-made	6	4	0	2
Others	11	0	1	2

Table 9: Detailed content of Table 8, for individual institutions.

5.1. Currently Used Database

5.1.1. Database

Table 8 represents currently used databases which the result for the question (1). The most frequently used one is ATR database comprising 27% of the total use followed by ASJ database's 20%. Not small number of institutions use their own self-made databases built from their own needs. "Other databases" refers to small databases that are not listed in this paper. The RWCP, ATR, ASJ databases include all datasets that belong to each category. If we look at the result more closely with individual institutions, it becomes Table 9 which is the detailed version of Table 8, where "Univ." stands for university, "Ind." for industry's research institutions, "Gov."

for government research institutions and “Unkn.” for the unknown (anonymous) group.

5.1.2. Purpose of Database Use

Databases that are currently used in many institutions have different nature when they were created first. The questionnaire respondents were analyzed according to their usage. The usage of speech database is summarized into the following five categories.

- (a) Speech recognition
- (b) Speech synthesis
- (c) Acoustic analysis
- (d) Sentence analysis
- (e) Speech/language education

The category (a) includes sentence comprehension, speech summarization, system development, evaluation, and speaker recognition, and (b) includes prosodic analysis while (c) involves speech coding research. The result

Database	(a)	(b)	(c)	(d)	(e)
JEIDA	12		1		
RWCP	5			2	
ATR	23	8	2		
ASJ	20	2	1	1	
TIMIT+TIDIGIT	5				1
CSJ	6		1	1	
Self-made	5	2	4	2	
Others	10	1	2		1
Total	86	13	11	6	2

Table 10: The purpose of database usage.

is listed in Table 10. The total number for individual database does not necessarily coincide with the total number listed in Table 8 since many datasets can be used in multi-purpose. It is clear that a majority of corpora are used in speech recognition and related research, about 73% of the total use, followed by synthesis research, but far less in frequency. Perhaps this result reflects the database-creating policy in our society.

5.2. Corpora for Future Speech Research

There is too little space left in this paper to introduce all the results of respondents. Let’s give a brief look at responses towards the question (5), about database that will be necessary/needed for future research work. The following is a list of top-ten corpora that the respondents think them important for future research.

1. Dialogue/continuous speech corpus recorded in real living environments (6 respondents).
2. Para-linguistic, emotional speech corpus (4)
3. Corpus for different speaking style (3)
4. Bi-modal, multi-modal speech corpus (2)
5. Read-speech corpus with prosody tag (2)
6. Corpus for broadcasting news speech data (2)
7. Corpus for speaker recognition including data for the same speakers recorded in different times and/or ages (2)
8. English corpus pronounced by the Japanese (2)
9. Mobile-phone speech data with a large number of speakers (1)

10. Speech corpus having various kinds of information about production including movie pictures (1)

There are many other responses about desirable future speech corpus that are excluded from the above list but we believe they are all important.

6. Acknowledgements

The authors would like to express their sincere gratitude to all institutions and personnel who took precious time to respond to the questionnaire and to those who gave us detailed information about their corpora.

7. References

- Campbell, N., 2001. The Recording of Emotional Speech – JST/CREST database research. *Proceedings of the Oriental COCOSA Workshop*, 78-80
- ESCA Workshop on Speech Input/Output Assessment and Speech Database, 1989, Noordwijkerhout, The Netherlands
- ESCA Workshop on International Coordination and Standardization of Speech Databases and Speech I/O Assessment Methods, 1991, Chiavari, Italy
- Furui, S., Maekawa, K., and Isahara, H., 2002. Intermediate results and perspectives of the project “Spontaneous Speech: Corpus and Processing Technology,” *Proceedings of the 2nd Spontaneous Speech Science and Technology Workshop*, 1-6
- ICSLP Workshop on International Coordination and Standardization of Speech Databases and Assessment Techniques for Speech Input/Output, 1990, Kobe, Japan
- Iso, K., et al, 1988. Design of a Japanese sentence list for a speech database. *Proceedings of Acoustic Society of Japan*, March
- Itahashi, S., 1985. Speech database of discrete words. *The Journal of Acoustic Society of Japan*, 41:723-726
- Itahashi, S., 1999. On recent speech corpora activities in Japan. *The Journal of Acoustic Society of Japan*, E20:163-169
- Kitazawa, S., et al, 2001. Preliminary study of Japanese MULTTEXT: a prosodic corpus. *Proceedings of ICSP*, 825-828
- Kuwabara, H., et al, 1989. Construction of a large-scale Japanese speech database and its management system. *Proceedings of ICASSP*, 560-563
- Kuwabara, H., 1996. Speech database of the priority area “Prosodic Features of Spoken Japanese.” *Humanities and Information Processing*, 12:60-61
- Maekawa, K., Koiso, H., Furui, S., and Isahara, H. 2000. Spontaneous speech corpus of Japan. *Proceedings of LREC Second International Conference*, 947-952
- Maekawa, K., 2002. Compilation of the corpus of spontaneous Japanese: A status report, *Proceedings of the 2nd Spontaneous Speech Science and Technology Workshop*, 7-10
- Nakamura, S., et al, 1999. Sound scene data collection in real acoustic environments. *J. of Acoust. Soc. Jpn. (E)*, 20:225-231
- Nakamura, S., et al, 2000. Acoustical sound database in real environments for sound scene understanding and hands-free speech recognition. *Proceedings of the 2nd International on Language Resources and Evaluation*
- Takeda, K., et al, 1988. A Japanese speech database for various kinds of research purposes. *The Journal of Acoustic Society of Japan*, 44:747-754