

A Task Analysis of Nursing Activities Using Spoken Corpora

Hiromi Itoh Ozaku¹, Akinori Abe¹, Kaoru Sagara²,
Noriaki Kuwahara¹ and Kiyoshi Kogure¹

¹ Advanced Telecommunications Research Institute International
{romi,ave,kuwahara,kogure}@atr.jp

² Seinan Jo Gakuin University
sagara@seinan-jo.ac.jp

Abstract. This paper illustrates our examination of the “E-nightingale Project,” reports the results manually obtained from task analyses of nursing activities using spoken corpora, and describes the possibility of automated task analyses using natural language processing technologies. Recently, medical malpractice has become a serious social problem and safety measures to prevent and decrease medical malpractice are being taken in some hospitals, for example, by building computerized database in new systems.

The Japanese Nursing Association states in its guidelines that nurses are encouraged to make nursing reports, including medical malpractice reports, and to analyze the cause of accidents, which is helpful for preventing recurrences.

However, it is very difficult for nurses to produce detailed records during their working hours except for malpractice reports. Furthermore, it is hard work for nurses on duty to analyze the recorded data in detail.

As a solution, we have launched the “E-nightingale Project,” in which some systems using wearable computers are being developed for the purpose of preventing and reducing medical malpractice.

As part of this project, we built spoken corpora using voice data that were monitored and recorded daily nursing assignments in hospitals with our developed devices. 800 hours of voice data were accumulated, and 70 hours of those were transcribed as the spoken corpora. Then we started analyzing nursing tasks using the spoken corpora, and considered the possibility of automated task analysis using natural language processing technologies.

1 Introduction

Recently, medical malpractice has become a serious social problem [1]. The Japanese Ministry of Health, Labor and Welfare has reported that nursing teams are most frequently involved in medical accidents in hospitals [2]. The Japanese Nursing Association also states in its guidelines that nurses are encouraged to make nursing reports and to analyze the cause of accidents by comparing medical malpractice reports, which is helpful for preventing recurrences. Some nursing

tasks in particular are closely associated with the occurrence of medical malpractice. In other words, if the nursing reports and medical malpractice reports are analyzed according to nursing tasks, it can be understood why and when a medical malpractice event happened. By using the results, medical malpractice can be reduced and prevented. However, it is very difficult for nurses to make detailed records during their working hours; furthermore, it is hard work for nurses on duty to analyze the recorded data in detail.

We launched the "E-nightingale Project" to develop a nursing service support system based on multimedia data in the real field that monitors and records nursing activities in detail using wearable computers. We can obtain multimedia data from our developed devices such as video data, sensed data of body actions, and voice data. Since video data can only be obtained in a narrow range, it is not possible to analyze all of the nursing activities. In fact, we are still analyzing sensed data so we cannot yet determine nursing behaviors from the sensed data. On the other hand, nurses have to confirm task names aloud every time they start their tasks, and we can use this information to analyze their tasks. Therefore, we focus on voice data to analyze nursing activities. In addition, since we need to develop dictionaries for an automatic speech recognition, we have decided to build corpora on nursing activities. This is the first attempt to develop special devices, to create spoken corpora for understanding nursing assignments, and to analyze nursing tasks by using the spoken corpora.

We monitored nurses in a hospital for several days, and recorded multimedia data using our wearable computers. The multimedia data include video data, voice data, sensor data and so on. We tried to analyze these data automatically; however, there arose some problems with analyzing them automatically. For example regarding voice data, nurses usually use technical terms with various expressions. We have some databases that can be regarded as standard dictionaries, such as JNPSM (Japan Nursing Practice Standard Master) built by MEDIS-DC (Medical Information Center Development Center) [3], and ICNP (International Classification for Nursing Practice) [4], but since they include written language, it is difficult to match their terms and the spoken terms collected in our project. Furthermore, concrete definitions of nursing tasks in actual working places do not exist in the research area of nursing task analysis.

In this paper, we report a way to accurately monitor nurses' activities in actual working places, and propose a method to effectively analyze nursing tasks from nurses' activity data. We first illustrate our examination of developing corpora of voice data for understanding nurses' activities, then report results from task analyses of nursing activities manually, and finally explain the possibility of automated task analyses using natural language processing technologies such as text-clustering methods.

2 E-nightingale Project

As previously mentioned, preventing medical malpractice is a very important research issue. To do so, it is vital to recognise the reality of hospital situations, and

to understand nursing activities. It is also important to make accurate nursing reports.

To fully understand nursing activities, a one-day schedule of nursing tasks should be recorded and analyzed. In the traditional method, nurses have reported their activities after their work using their memories. Alternatively, one researcher has followed a nurse all day and taken meticulous notes of the nurse's activities. The traditional method, called a "Time Study Survey," is expensive and cannot fully describe their daily activities with sufficient accuracy [7], which is why we started the E-nightingale Project. The project has the following purposes:

- to develop wearable computers for easy recording of daily nursing assignments,
- to obtain multimedia data of nursing assignments in the hospital,
- to analyze the data to accurately grasp nursing tasks and to build a database of nursing tasks efficiently,
- to utilize the database by better understanding the activities and comparing malpractice reports to prevent future malpractice.

Though we can obtain multiple data types, in this paper we focus on voice data for the following reasons:

- natural language processing technologies are evolving and can be applied in the real field,
- speech recognition technologies are also evolving and can be applied in noisy fields [5, 6],
- much text information can be utilized by the diffusion of electronic charts,
- nursing activities and situations can be understood easily if the speech of nurses can be grasped.

This is to say we are developing a method and system for automatic "Time Study Survey." As the first step, we recorded the voice data of nursing activities along with cue words and cue sentences such as "I will start A task," that are regarded as annotations by nurses using our devices. In the next section, we explain how to obtain the voice data.

3 Experiments to Collect Data

We recorded voice data in some hospitals using our wearable computer. The voice data contained daily nursing assignments with many types of sounds, including talks with patients, and/or doctors, noise during nursing care, and other tasks.

First, we explain the wearable computer used in this investigation. To obtain voice data in hospitals, we used a special device comprising of an ICrecorder,

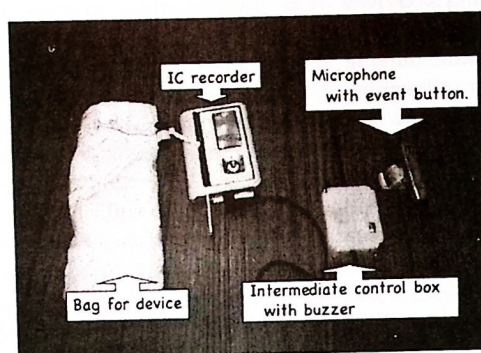


Fig. 1. Our Devices

a microphone with an event button, and an intermediate control box with a buzzer, as shown in Fig. 1.

The event button is used for explicit voice annotation when nurses start or complete a task. Recently, to decrease medical malpractice, nurses have had to confirm a medicine's name aloud when they start mixing medicines into an intravenous drip. If the voice is recorded by our device, we can correlate the voice with the nursing task "Medication Administration" manually. Furthermore, if conversations between nurses and patients can be recorded, the task performed at that time can be easily understood.

When the button is pushed, the buzzer sounds once and its sound is recorded, and then nurses record their tasks of the moment by speaking short sentences that include words of confirmation and conversations with patients. The buzzer is also set to sound periodically (every 10 minutes) to prompt nurses to make voice inputs about their ongoing tasks. Simple signal processing can extract and classify task-driven and periodic voice records as well as nurse call rings.

In two departments of a hospital, we conducted experiments on collecting data of nursing tasks through voice annotation. The nurses worked in three shifts, assigned to the primary patients of each ward. All nurses were given instructions on how to use our devices. The entire recording time was about 800 hours. Data were gathered from 163 trials for a 14-day experiment involving 39 nurses using our devices.

Next, we transcribed the voice data to text to build nursing spoken corpora — the Nursing Task Corpus and the Nursing Dialogue Corpus. The following section will include a detailed explanation of the corpora.

4 Nursing Spoken Corpus

We transcribed the voice data to text to build the nursing spoken corpora that are used for understanding nurses' tasks and collecting nursing technical terms in the daily nurse assignments, especially in clinical meetings.

The process of creating the corpora is as follows:

1. Data of the sounds, including buzzer and short sentences, are broached by the signal processing.
2. Transcriptions of each broached data are made manually.
3. Task names in a job category list are manually attached as tags to each transcription of short sentences for the Nursing Task Corpora.
4. For the Nursing Dialogue Corpora, words not included in the IPA dictionary [8] but included in voice data in conferences or clinical meetings are transcribed and tagged as "unknown words."

The transcription was made by four staff members, including an experienced specialist in making transcriptions, a nurse who had worked in hospitals for three years or longer, a pharmacist, and a part-time employee, and it was assumed that the latter three had no experience working in the field of transcription.

Two types of corpora were built from the transcribed data. One was the Nursing Task Corpus, the other the Nursing Dialogue Corpus. The Nursing Task Corpus includes transcriptions of voice data for 10 seconds with a buzzer that were extracted by simple signal processing from the voice data of a one-day assignment. The Nursing Task Corpus includes all activities by a nurse in one day. This is useful for making accurate nursing reports. If the Nursing Task Corpus is analyzed together with medical malpractice reports in detail, nurses can find when malpractice events happen, what nurses do, and how such events occur and so on. The Nursing Dialogue Corpus includes all conversations during clinical meetings. We can thus easily understand the communications among nurses, and collect technical terms in the real field for automatic speech recognition from the Nursing Dialogue Corpus.

In the next section, we briefly explain the Nursing Dialogue and Nursing Task Corpora.

4.1 Nursing Dialogue Corpus

During nurses' shift changes, they hold clinical meetings to discuss patient information, and in the process, they modify and confirm this information. Furthermore, if problems occur in their work area, they hold brief conferences to solve them by discussing their experiences.

We think dialogue in the clinical meetings includes cue words of nursing activities, their work flows, nursing technical words, and so on, and consider the cue words to be very useful for understanding their tasks. In addition, it is also useful as a language model to transcribe the voice data automatically by using speech recognition tools. We collected technical terms from dialogues in clinical meetings, especially medicine names, and medical test names. Because these names are treated as "unknown words," they do not exist in commonly used dictionaries. An example from the corpus is shown in Table 1. In this case,

"ENTO"³ is an unknown word and should be treated as a cue word. Because "ENTO" helps to confirm nursing activities [9].

Table 1. Example from the Nursing Dialogue Corpus

Time	Nurse-ID	Utterance
00:50:19	A	Kogure-san will be discharged tomorrow. Please do not forget the ENTO sheet.
00:50:20	B	Yeah, .. ENTO sheet..
00:50:40	C	The drip infusion is still being given, isn't it?
00:50:42	A	Yes. It's still dripping now.
00:50:44	B	Ah, until when?
00:50:45	A	Until tomorrow. Until 6 o'clock.

Supporting communication between doctors and patients is vital when doctors give informed consent. As such, some researchers have implemented special communication support systems tailored for communication between doctors and patients [10]. We think our corpora are helpful for clarifying communication mechanism in hospitals.

However, the nurses' dialogues have not only a complex communication mechanism, but also many expressions such as nursing technical terms. For instance, in general "せんがん (sengan)" means "洗顔 (washing face)," but if it is narrated before surgery in ophthalmology, it means "洗眼 (washing eye)." "自立する (jiritsu-suru)" means "standing walk" if it is narrated in a conversation among nurses, but it means "to earn one's living by oneself" in general, as when it is narrated in a conversation between a nurse and a patient's family. Thus some words have multiple meanings according to the situation. As a result, sometimes nurses, especially novice nurses, experience misunderstanding in their communications. Another example is "ほりゅう (horyu)," which means "suspend giving medicine" or "keep a sting stung." Therefore, we think it will be possible to gather many words that have multiple meanings when we check the corpus. This will allow us to build or extend the contents of a dictionary of multiple-meaning expressions. We consider these expressions to include cue words that are nursing technical terms, and as such they are useful as language models to automatically recognize speech and to prevent malpractices due to misunderstanding.

4.2 Nursing Task Corpus

For the Nursing Task Corpus, when nurses started or completed a task, they pushed the event button and recorded their tasks of the moment by speaking short sentences. These sentences include confirmation messages for nursing

³ "ENTO" means "to discharge from hospital," originating from the German word "Entlassen."

tasks, conversations with patients, and so on. First, we transcribed these short sentences and related them to nursing tasks associated with each one. To effectively build the Nursing Task Corpora, we also annotated job categories for each nursing task. We used terms for job categories included in Kango Gyoumu Shishin (看護業務指針) published by the Japanese Nursing Association (日本看護協会) [11] and Classification of Nursing Practice (看護行為用語分類) published by the Japan Academy of Nursing Science (日本看護科学学会) [12]. These annotations were tagged by one ex-nurse who had worked in hospitals for three years or longer, just like for making transcriptions. An example from the corpora is shown in Table 2.

Table 2. Example from the Nursing Task Corpora

Time	Utterance	Job Category
11:01:00	I'm going to join a short conference.	18-106 conference
11:20:48	The short conference is finished.	18-106 conference
11:28:11	I'm going to prepare a drip infusion set for Abe-san.	13-63-6A0502 intravenous infusion
11:32:01	I've finished preparing the drip for Abe-san.	13-63-6A0502 intravenous infusion

In this paper, since we focus on task analysis, we will discuss task analysis using the Nursing Task Corpus in the following sections.

5 Results and Discussion of Task Analysis

In this section, we show a task analysis using the spoken corpora, consider the results, and discuss our study in relation to automatic task analysis.

5.1 Term Definition for Task Analysis

To make a task corpus such as that shown in Table 2, we used task names (task terms), explained previously, as job categories. Table 3 shows an example of the job category list. It is difficult to find direct relations among these categories and nursing task corpora in the real field because they have lexical ambiguity, various expressions, and are too detailed for analysis of the relation between nursing tasks and medical malpractice. We defined some task terms for the nursing assignments, and correlated the task terms with the job category list.

The task terms are defined as follows: One-day nursing jobs, called **Daily Nursing Assignments**, are divided into many tasks that are determined by the job categories list etc. Daily Nursing Assignments consist of many **Nursing Tasks**, for example, assistance with meals, instillation of drips, preparation of infusions, and so on. Nursing Tasks are divided further into two categories, i.e.,

Table 3. A Part of the Job Category List

No	Job List 1 (Major Division)	No	Job List 2	Cord No	Term of CNPBook (Minor Division)
1	Hygiene (身体の清潔)	1	Bed-bath (清拭)	2D0301	Hand-bath (手浴)
				2D0302	Foot-bath (足浴)
		2	Hair-care (洗髪, 整髪)	2D0601	Body-bath (全身清拭)
				2D0401	Washinghair (洗髪)
				2D0301	Stylinghair (整髪, 結髪)
6	Patient mover (患者の移送)	26	Patient mover	2G0201	Move to wheelchair (車椅子への移乗)
				2G0301	Move with assistive device (補助器具を用いた移動の介助)
				2G0401	Other transfer (移送)
13	Doctor's assistant (診療, 治療の介助)	58	Rounds (回診)		
		59	Change bandage (包交)	6C0101	Injurycare (創傷ケア)
		60	Brace (ギブス)		
		63	IVH-care (IVH, 持続点滴の管理)	6A0502	Intravenous infusion (点滴静脈内注射)
18	Takeover among nurses (看護師間の報告, 申し送り)	105	Changeover (申し送り)		
		106	Conference (カンファレンス)		
34	Control of staff' health care (職員の健康管理)	162	Meal (食事)		
		163	Rest, Break Time (休息, 休憩)		

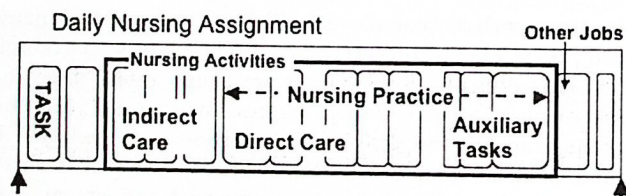


Fig. 2. A model of Term Definition of Nursing Tasks

Nursing Activities and Other Tasks. Nursing Activities include jobs that nurses have to do, and are again divided into two categories, i.e., **Nursing Practices** and **Indirect Care**. Nursing Practices include jobs directly related to patients, consist of **Direct Care** and **Auxiliary Tasks** that support doctors. A standardized model of the term definition of nursing tasks is shown in Fig. 2.

The relations of our term definition and the job categories list are shown in Table 4.

Table 4. Relation of the Term Definition and the Job Categories List

Term Definition	No of Job Category List	Job example
Direct Care	1 – 12	Bedside-bath, Care of hospitalization and release
Auxiliary Tasks	13 – 16	Doctor assistance, Vital care, Medical tests
Indirect Care	17 – 27	Report of medical condition, Conference among nurses
Other Tasks	28 – 35	Messenger tasks, Communication with other sections

5.2 Results of Task Analysis from Nursing Task Corpora

We created the Nursing Task Corpora to analyze nursing tasks for the whole day. We annotated job categories for each transcribed sentence, analyzed the working time taken for the four categories (Direct Care, Auxiliary Tasks, Indirect Care, and Other Tasks) using the Nursing Task corpora. Example results are shown in Table 5. These results are average values of task analysis using all data in each shift in one department.

Indirect Care in the Nursing Activities, such as takeover among nurses, and preparing medicines, takes a long time in all shifts. The time for Direct Care and Auxiliary Tasks is unexpectedly short in all shifts. Therefore, we think we can provide some practical methods to concentrate attention on only the time for Direct Care and Auxiliary Tasks.

Table 5. Example Results of Job Analysis using the Nursing Task Corpora

Job Categories	Dayshift	Twilightshift	Nightshift
Direct Care	0:59:06	0:54:12	1:36:27
Auxiliary Tasks	1:40:19	1:26:28	1:53:22
Indirect Care	2:52:11	2:10:53	3:15:50
Other Tasks	0:53:24	0:16:53	1:05:45
Unspecified	1:30:08	0:50:14	1:23:40
Total Time	7:55:08	5:38:39	9:15:04

5.3 Discussion of Task Analysis

The time periods for Direct Care and Auxiliary Tasks are very important ones for preventing medical malpractice because nurses give medical treatment to patients directly. If nurses can focus their attention on these time zones, we think there will be great success in preventing medical malpractice.

To realize this, we need to know which tasks nurses are doing as they do them. This can be done by making transcriptions with the aid of speech recognition tools [13-15]. For the next step, we need methods to analyze nursing tasks automatically. One solution could be applying text classification technologies to task analysis; that is, if transcribed sentences are classified into the same job category, the sentences will include at least one of the same terms. We investigated the features of the task corpora in detail to explore the possibility of applying text classification to the nursing task analysis. We checked about 6,000 utterances in a part of the task corpora as a trial. Thirty-nine nurses participated in the trial experiment. They were divided into 35 job categories within the major divisions⁴, such as 1-1-2D0401 (Hygiene/Hair-care/Washing-hair 洗髪), 13-58 (Doctor assistance/Rounds 回診), as shown in Table 3. All transcriptions in the trial data were analyzed by the morphological analysis tool Chasen [16]. The frequency of nouns in the transcriptions was 18,737 times (1,590 words), and the frequency of verbs was 6,321 times (530 words). In fact, one utterance includes one verb and two or three nouns, suggesting that each sentence is short because of spoken language.

In the data, the most frequently occurring verb is し (do). With 1,400 appearances this word appears in every transcription of every job category. Verbs, however, are of limited use for understanding nursing tasks. On the other hand, the most frequently occurring noun is さん (Mr, Miss, Mrs.), occurring 1,948 times. This is also useless for understanding the tasks, but nouns of high frequency are terms that are easy to associate with the Job Categories, such as チェック (Check), カルテ (Medical chart), 引き継ぎ (Takeover). We checked the relation between high-frequency nouns and job categories. The features are shown in Table 6.

We randomly checked the features of each shift in one department, the frequent appearance of nouns divided in the same categories. This means that the surface co-occurrence relations between nouns of utterances and job categories are useful for task analysis. Though our corpora include very few data for one department in a certain hospital for analyzing general nursing tasks, we think it is possible to understand nursing tasks automatically using the surface co-occurrence relations between nouns and job categories.

6 Conclusion

We introduced our project and explained the process of task analysis using the nursing task corpora and showed how to analyze daily nursing assignments efficiently using a wearable computer. We have now obtained 800 hours of voice data

⁴ There are about 460 job categories in the minor division.

Table 6. Features of Relations with High-Frequency Nouns and Job Categories

Noun	Frequency	Job Categories	Appearance Rate %
チェック (Check)	410	18-109 情報の整理 (Summary of medical chart)	58.2 (239/410)
カルテ (Medical chart)	353	18-109 情報の整理 (Summary of medical chart)	68.6 (242/353)
引き継ぎ (Takeover)	212	18-105 申し送り (Changeover)	89.6 (190/212)
点滴 (Drip)	254	13-63-6A0502 点滴静脈内注射 (Intravenous infusion)	68.5 (174/254)
検温 (Temperature check)	171	15-74 バイタルチェック (Vital check)	84.8 (145/171)

to produce a 20-hour task corpus and a 50-hour dialogue corpus. As a first step, we have been manually transcribing the spoken corpora, and classifying the job categories. Furthermore, we found a correlation between high-frequency terms and job categories. If we can build or extend the contents of a concept-based, a multiple-expression dictionary, and an abbreviated-expression dictionary, we can produce additional corpora and analyze nursing tasks automatically using technologies of speech recognition, morphological analysis, and classification.

Utilizing the spoken corpora, we can expect to understand nursing assignments in detail and the structure of communication among nurses. As a result, it should be possible to prevent medical malpractices due to miscommunication, and to call nurses' attention to certain tasks correlated with medical malpractice. We think a method that makes spoken corpora can produce nursing reports more easily and accurately than "Time Study Survey" can. In future work we will clarify the method in detail and apply it to many hospitals.

In this paper, we only analyzed data from one department in a hospital. However, we have plans to obtain data from some other departments using our devices. Furthermore, by incorporating detailed spoken corpora with other multiple data, we will pursue the exact cause of medical malpractice in hospitals and undertake work to build a nursing ontology.

Acknowledgments

We would like to thank the nurses for their cooperation in our experiment. This research is funded by the National Institute of Information and Communications Technology.

References

1. L. T. Kohn, J. M. Corrigan, and M.S. Donaldson: "To Err Is Human: Building a Safer Health System." The National Academies Press. (1999).

2. Healthcare Safety Precaution Network Project:
["http://www.mhlw.go.jp/topics/2001/0110/to1030-1.html#2-1"](http://www.mhlw.go.jp/topics/2001/0110/to1030-1.html#2-1)
in Japanese, (2001).
3. Medical Information Center Development Center eds.:
"Japan Nursing Practice Standard Master"
http://www.medis.or.jp/4_hyojyun/kango/kango.html in Japanese.
4. Japanese Nursing Association Nursing Practice International Classification Research Project translated,
"International Classification of Nursing Practice (Japanese version)," <http://icnp.umin.jp/> in Japanese.
5. M. J. F. Gales and S. J. Young: "Robust Continuous Speech Recognition Using Parallel Model Combination," IEEE Transactions on Speech and Audio Processing, Vol. 4, No. 5, pp. 352-359, May 1996.
6. S. Nakamura, K. Takeda, T. Yamada, S. Kuroiwa, T. Nishiura, M. Misumachi, 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, pp. 535-544, 2005.
7. N. Kuwahara, K. Kogure, N. Hagita, H. Iseki: "Ubiquitous and Wearable Sensing for Monitoring Nurses' Activities." The 8th World Multiconference on Systemics, Cybernetics and Informatics SCI 2004.
8. "IPADIC version 2.7.0 (IPA 品詞体系辞書)" <http://chasen.aist-nara.ac.jp/stable/ipadic/ipadic-2.7.0.tar.gz>
9. H. Ozaku, A. Abe, N. Kuwahara, F. Naya, K. Kogure and K. Sagara: "Building Dialogue Corpora for Nursing Activity Analysis" In the proceedings of Sixth International Workshop on Linguistically Interpreted Corpora, (2005).
10. K. Katsuyama, Y. Kouyama, Y. Hirano, K. Mase, and K. Yamauchi: "Doctor-Patient Communication Support Method by Visualizing Topic Structure," In the Japan Journal of Medical Informatics, pp. 578-587, Vol. 24, No. 6, 2004. (in Japanese.)
11. "Kango Gyoumu Kubun Hyou (看護業務区分表)," Kango Gyoumu Kijunshu (看護業務基準集), pp. 336-337, published by Japan Nursing Association (日本看護協会). ISBN 4-8180-1089-8, 2004.
12. "Classification of Nursing Practice (看護行為用語分類)," Japan Academy of Nursing Science (日本看護科学学会). ISBN 4-8180-1142-8, 2005.
13. S. Nakamura: "Towards Robust Speech Recognition to Acoustic Disturbances." In the Journal of the Acoustical Society of Japan, pp. 662-667, Vol. 57, No. 10, 2001. (in Japanese.)
14. Julius - open source real-time large vocabulary speech recognition engine: ["http://julius.sourceforge.jp/en/index.html"](http://julius.sourceforge.jp/en/index.html)
15. M. Fujimoto and S. Nakamura: "Particle Filtering and Polyak Averaging-based Non-stationary Noise Tracking for ASR in Noise," In the Proceedings of IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU05), pp. 337-342, Nov. 2005.
16. Chasen 2.3.3 ["http://chasen.naist.jp/hiki.Chasen/"](http://chasen.naist.jp/hiki.Chasen/)