



# GEMSIS — a novel application of speech recognition to emergency and disaster medicine

Satoshi Tamura<sup>1</sup>, Kunihiro Takamatsu<sup>2</sup>, Shinji Ogura<sup>2</sup>, Satoru Hayamizu<sup>1</sup>

<sup>1</sup> Department of Information Science, Gifu University, Japan

<sup>2</sup> Department of Emergency & Disaster Medicine,  
Gifu University Graduate School of Medicine, Japan

E-mail: {tamura@info., gaosong@, oguras@, hayamizu}@gifu-u.ac.jp

## Abstract

This paper reports an instance of novel application of speech recognition applied to emergency and disaster medicine. The emergency medical system named “GEMSIS” (Gifu Emergency Medical Supporting Intelligent System) including the speech recognition application is also introduced in this paper. Speech recognition plays an important role in this system; when a paramedic team is sent to a disaster or accident site, a life-saving technician reports the situation using speech recognition in the site. The recognized information are shared by all hospitals and critical care centers. This system can solve the severe issue of the emergency medical care in which pre-hospital medical care is insufficient due to lack of information. A prototype application of speech recognition interface was constructed to evaluate a baseline performance and to make a discussion with medical doctors. Through this work, it is found that the applicable domain of speech processing technology can be extended.

**Index Terms:** speech recognition, emergency medicine, information share, medical-engineering research.

## 1. Introduction

First of all, in this paper, we develop the information share system named “GEMSIS” for emergency medical care, including speech recognition technology [1]. Backgrounds of the system are introduced below.

For several decades, speech recognition technology has been investigated to enhance the performance and to be driven in real time. It contributes recent speech recognition applications, such as spoken dialog system, dictation [2], car navigation [3], and other state-of-the-art applications [4]. However, unfortunately people often give up using speech recognition in these systems. Due to background noises, uncertain utterances, and unexpected sentences, the performance of speech recognition drastically degrades. It makes the user think that speech recognition is no more usable. Another reason why people do not use speech recognition is other alternatives are available in many cases. People can input information to a system by keyboard and mouse, or telephone. There is the excellent engineering “seed” named speech recognition, however, much appropriate “fields” for the seed have not been found.

On the other hand, emergency medicine has aimed at improvement of the life-saving ratio: the ratio of the patients who are alive one month later to all emergency patients. There are several issues regarding emergency transportation; in an accident scene, a life-saving technician or a paramedic can hardly select the optimal hospital according to the patient’s trauma. An emergency hospital sometimes cannot make the preparation to

accept a patient since a doctor does not know the patient’s information until the patient arrives. The disaster medicine also has a problem; the present emergency cooperation among the hospitals is insufficient to a large disaster such as an earthquake (e.g. the Great Hanshin earthquake in 1995 where over 50,000 people were injured or killed [5]) or a train/airplane accident (e.g. the Amagasaki rail crash in 2005 where roughly 700 passengers were hurt or lost [6]), where many people are simultaneously injured throughout the disaster area. The system which enables real-time information share of patients and medical resources of emergency hospitals is thereby required. However, there is the severe problem how a technician inputs patient’s information to the system in a hands-busy situation with conducting resuscitation. There is a “need” to handsfree information input method in the emergency medical field.

We innovatively unite the engineering “seed” with the medical “need”. Then as already described, the information share intelligent system for emergency medicine using speech recognition has been developed; in our system, a life-saving crew inputs patient’s information using wireless microphone, and speech recognition is conducted to convert the utterance to a text. The text is transferred to the system server to be shared by doctors in hospitals. Our system provides large impact and great significance not only to emergency medicine but also to speech processing technology. By using our system, pre-hospital medicine varies and improves revolutionarily, resulting the increase of quality of emergency medical care and the decrease of preventable trauma death. Regarding speech processing, our system becomes a leading case to expand the domain of speech recognition.

This paper introduces the concept of our GEMSIS project, as well as the speech recognition interface of the system. A test model of speech recognition device was developed and preliminary experiment was investigated.

This paper is organized as follows: in Section 2, the summary of GEMSIS is introduced. Section 3 describes the speech interface in GEMSIS, and experiment and demonstration are performed in Section 4. Finally Section 5 concludes this paper.

## 2. GEMSIS overview

A GEMSIS is the multipurpose emergency medical supporting system based on cooperative distributed network model. The system accomplishes the wide-area cooperation among emergency medical organizations such as hospitals, fire stations, rescue teams, and local governments. Although the GEMSIS has several aims, in this paper we focus on the major purpose of the GEMSIS: a support for mass-gathering medicine and emer-

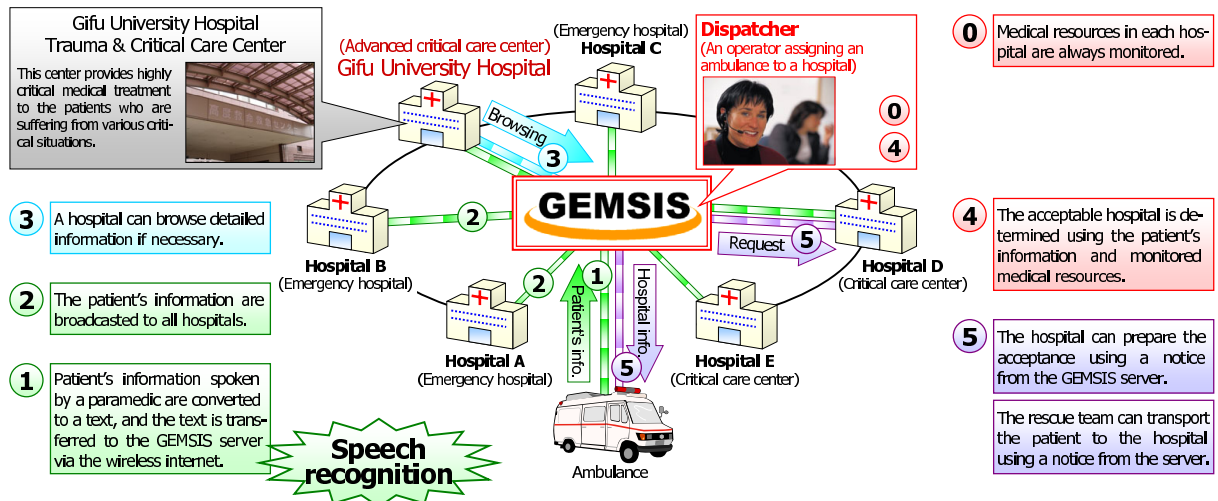


Figure 1: An overview of information share and dispatching an ambulance in our GEMSIS.

gency activities in a wide-area natural disaster or a massive accident.

The GEMSIS produces a function of information share in emergency medicine and an effective dispatching method of emergency transportation. Figure 1 indicates the overview.

- (0) Medical resources (the number of vacant beds and available doctors in every emergency hospitals and ambulance assignment) are monitored constantly by the GEMSIS server. Anyone working at hospital can browse present resources.
- (1) If an ambulance is sent to an accident site, a life-saving technician inputs patient's information in the site by uttering them to a speech recognizer. The speech data and the recognized text are wirelessly transmitted to the GEMSIS server as a first call.
- (2) A summary of the first call is broadcasted to all emergency hospitals and critical care center.
- (3) Any doctor in the hospital or the center can refer to the detailed patient's and accident information.
- (4) A dispatcher responsibly determines which hospital or center is acceptable according to the first call and medical resources. The GEMSIS also supports its decision using some intelligent architectures.
- (5) The GEMSIS server sends a request of acceptance to the selected hospital. The hospital can prepare the acceptance before the patient arrives. The server also sends the result to the paramedic team, enabling them to transport their patient to possibly optimized hospital.

### 3. Speech recognition interface

#### 3.1. Why is speech recognition essential in GEMSIS ?

In the scheme described in Section 2, it is necessary to input patient's information to the server. However, an emergency life-saving technician cannot use both hands due to life-saving emergency treatment to the patient, and there is no time for the technician to input the information. In conclusion, inputting the data by keyboard or mouse cannot be implemented. Speech recognition can solve this problem. The technician can input the patient's information using speech recognition without hands.

There are a lot of advantages for speech recognition in this system. Since the life-saving technicians are trained to inform scar circumstance to a doctor clearly and precisely, their utterances are easy to recognize. It is possible to build the acoustic model adapted to each technician. These contribute improvement of recognition performance. By using a head set with a microphone, effect of background noise can be reduced. The report of emergency treatment and scar circumstance of the patient is made based on the JPTEC (Japan Pre-hospital Trauma Evaluation and Care) guideline [7]. The JPTEC is the pre-hospital trauma educational program, similar to the PH-TLS (Pre-Hospital Trauma Life Support) proposed in United States [8]. The JPTEC thus strongly restricts the technician's speech; most utterances spoken by the technician consist of a fixed phrase, and variations of sentence patterns are limited. Using a word-network-based language model, the contents of the utterance can be well estimated. Speech signals are transmitted wirelessly from the microphone to a PC or mobile device where speech recognition is conducted. These wireless instruments do not prevent relief activities of the technician.

#### 3.2. System overview

Figure 2 illustrates the overview of speech processing in information share of emergency transportation. In an accident site, a life-saving technician utters patient's information and condition: e.g. patient's gender and age, injured site, neurological scale, whether the patient is CPA (CardioPulmonary Arrest) or not. The utterances are recorded in a microphone of a Bluetooth<sup>TM</sup> head set. From the head set, speech signals are wirelessly transferred to a PDA (Personal Digital Assistant) with the technician. In the PDA, speech recognition is conducted using adapted acoustic models to obtain a reliable recognized text. Next, the downsampled speech signals and recognition results are transferred to a relay server in an ambulance through an ad-hoc network. Added ambulance information, e.g. its global position obtained from the GPS, the speech data including recognition text are finally uploaded to the GEMSIS server. Since the internet and wireless network are used, the Secure Socket Layer (SSL) protocol is implemented.

The recognized text might include recognition errors. In

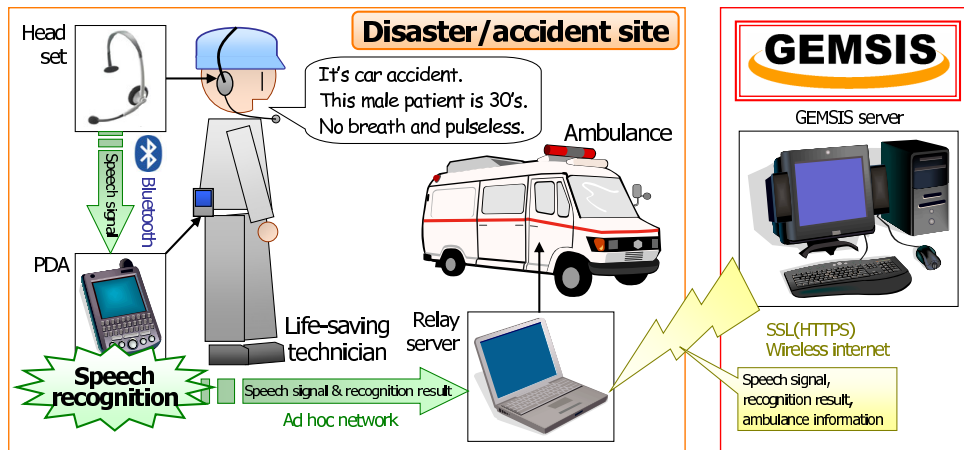


Figure 2: An overview of speech recognition in our GEMISIS.

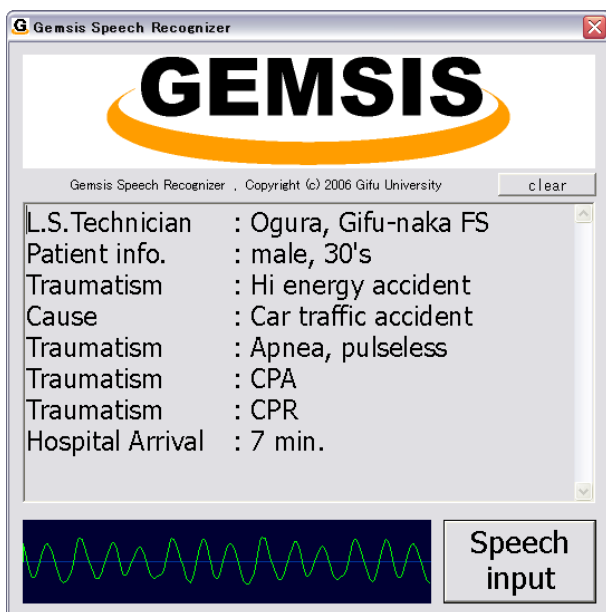


Figure 3: A PDA application for life-saving technician.

the GEMISIS, some methods are assumed to correct these errors. Knowledge-based processing is one of the solutions, where contradictory items are rejected or only one confidential item is adopted. Alternatively, using the speech signals, the dispatcher can correct the errors. The speech data also help doctors to know the details of the first call that are not represented in the recognition texts.

### 3.3. PDA application

The interface application for a life-saving technician executed on a PDA device is developed. Figure 3 illustrates the interface of the application. Speech recording begins by clicking the "Speech input" button. Recorded speech waveforms are shown in the bottom of the application. Speech recognition is simultaneously executed, and then sentence kind and keywords are obtained and displayed from a recognized sentence.

### 3.4. Speech recognition function

Figure 4 illustrates a flowchart of speech processing. The task is sentence kind estimation and keyword detection, which es-

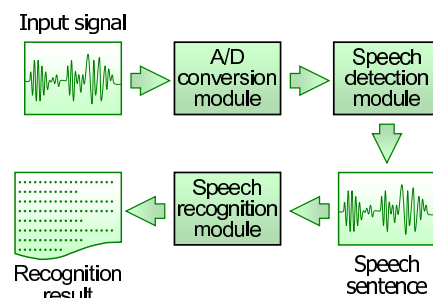


Figure 4: A flowchart of speech recognition.

Table 1: Speech recognition specifications.

Sampling frequency	16kHz
Quantization bit number	16
Frame length	25ms
Frame period	10ms
Acoustic features	MFCC (12 dim.), $\Delta$ MFCC (12 dim.), $\Delta$ power (1 dim.)

timated vocabulary size is roughly up to 200. Conventional speech recognition technology is used in our system. Speech recognition consists of the following modules.

#### 3.4.1. A/D conversion modules

Input audio signals are sampled and converted to digital data in this module. The specification of A/D conversion is shown in Table 1.

#### 3.4.2. Speech detection module

In this module, an energy coefficient is obtained at every intervals. A speech segment is detected by finding the region where the energy values are always superior to the threshold. To avoid recognition error, a speech segment which is shorter than one second is discarded. Consequently, the detected speech segment is extracted from the audio as a sentence speech data.

#### 3.4.3. Speech recognition module

Acoustic features are obtained from each frame in the speech segment. Speech recognition is subsequently conducted using

Table 2: An example of utterance of life-saving technician.

This is Ogura of the Gifu-naka fire station. The patient is 30's, and the gender is male. It is a high energy accident. Well, a car accident. There is no breath and pulse beat. He is now CPA (CardioPulmonary Arrest). CPR (CardioPulmonary Resuscitation) is now conducted. We will arrive at the hospital approximately 7 minutes later.

Table 3: Acceptable sentence kinds and keywords in a demonstration application.

Sentence kind	Technician's information, Patient age and gender, Traumatism, Cause, Hospital arrival time estimation
keyword (example)	Ogura, Yamada, Gifu-naka FS, 10's, 20's, ..., 90's, male, female, Hi energy accident, CPR, CPA, Apnea, Breath, Pulseless, Pulse beat, Car accident, Bicycle accident, 3min., 4min., ..., 10min.

the Viterbi algorithm, and the recognition result is obtained. A sentence kind and keywords are extracted from the recognized text. The sentence kind includes technician's information, patient's information, traumatism, and so on.

## 4. Experiment and demonstration

### 4.1. Demonstration application

In order to investigate the effectiveness, performance and impact of our system, a demonstration application was developed. This demonstration application included the interface shown in Figure 3 and the speech recognition function described in Section 3.4. This application was able to be executed in a note PC. The speech recognition module was constructed based on the HTK (HMM Tool Kit). Context-independent triphone HMMs, which were trained using the CSJ corpus [9], were adopted so as to easily register a new word. A word network was used as a language model in this application, where the subset of JPTEC was implemented; five sentence kinds and 43 keywords could be recognized. An example of utterances as well as acceptable sentence kinds and keywords are indicated in Table 2 and 3 respectively.

### 4.2. Operation verification

The verification test was conducted to examine a baseline performance. One male speaker uttered several evaluation sentences in clean environment using the demonstration application. Neither acoustic model adaptation nor cepstral mean normalization was conducted. The result shows over 90% keyword detection correctness. Some recognition errors were observed: acoustically similar abbreviations such as "CPR" and "CPA", and syntactically ambiguous sentences, e.g. a recognition error "Apnea, pulse beat" to a sentence "No breath and pulse beat" corresponding to the correct text "Apnea, pulseless".

### 4.3. Hearing and demonstration

The demonstration was conducted in several emergency medical domestic conferences, symposiums, seminars and meetings. After the demonstration, hearing and discussion were conducted. Since in Japan there are several severe earthquakes per year, many attendees were interested in our work. Comprehensively, their opinions and discussions show that our speech interface of the GEMSIS is much advanced, creative and needful.

## 5. Conclusions

In this paper, we introduced the information share system "GEMSIS" for emergency medicine, focusing on speech recognition interface. A paramedic crew can inform patient's information to hospitals by using speech recognition technology, in the situation where a handsfree method is needed. Speech recognition in our system thus assists pre-hospital emergency care effectively. Since the cooperation of emergency medical organizations is strongly required, the implementation of the GEMSIS has great social significance. Additionally, in this paper, speech recognition was applied to the field where it has not been used. And consequently, positive opinions and agreements were obtained. Then this work has another significance to extend the applicable domain of speech recognition. Finally, it is turned out that the medical-engineering cooperation has a possibility to accomplish an innovative result.

Our future work includes: (1) development of a full-scale application, (2) increase of the robustness of speech recognition through a field test, (3) cooperation with other functions in the GEMSIS, e.g. intelligent architecture, and (4) investigation of other devices using speech processing technology for emergency medicine.

## 6. References

- [1] K.Takamatsu et al., "GEMSIS, an introduction of intelligent information support system for emergency and disaster medicine," Proc. The 21st pacific science congress (2007-6, to be published).
- [2] T.Kawahara et al., "Free software toolkit for Japanese large vocabulary continuous speech recognition," Proc. IC-SLP2000, vol.4, pp.476-479 (2000-10).
- [3] M.Westphal and A.Waibel, "Towards spontaneous speech recognition for on-board car navigation and information systems," Proc. Eurospeech'99, vol.5, pp.1955-1958 (1999-9).
- [4] S.Tamura et al., "Automatic metadata generation and video editing based on speech and image recognition for medical education contents," Proc. INTERSPEECH2006, pp.2466-2469 (2006-9).
- [5] T.Ukai, "The great Hanshin-Awaji earthquake and the problems with emergency medical care," Renal failure, Marcel Dekker, New York, vol.19, no.5, pp.633-645 (1997-9).
- [6] T.Nagata et al., "Express railway disaster in Amagasaki: a review of urban disaster response capacity in Japan," Journal of Prehospital and Disaster Medicine, vol.21, no.5, pp.345-352 (2006-10).
- [7] <http://www.jp-tec.jp/> (written in Japanese).
- [8] <http://www.naemt.org/PHTLS/>.
- [9] K.Maekawa et al., "Spontaneous speech corpus of Japanese," Proc. 2nd conf. on Language Resources and Evaluation, pp.947-952 (2000-6).