Development of information systems in Japan Broadcasting Corporation

Eiichi Miyasaka
Faculty of Environmental and Information Studies
Musashi Institute of Technology

3-3-1, Ushikubo-Nishi, Tsuzuki-ku, Yokohana city, 224-0015, Kanagawa, JAPAN

Abstract

This paper describes some solutions for the problems which may occur in the process of introducing experimental systems developed in technical laboratories in the broadcasting news fields. An example presented here is the automatic captioning system using the automatic speech recognition system developed by a study group at NHK (Japan Broadcasting Corporation) Science & Technical Research Laboratories. One of the most important issues to be settled is that the developers should design the prototype, considering the users' potential fears for the new systems, including the consideration of potential "bugs" that may occur within the system. It will also be important for the developers to obtain the trust of the users in the fields before installation.

Keywords

information system, automatic speech recognition system, real-time transcription system, closed caption, decision flow

1. Introduction

Japanese digital HDTV satellite broadcasting was launched in December 2000, generating new types of services such as datacasting and interactive services as well as seven channels presenting HDTV programs.

Datacasting services bring new menus on a screen where viewers can select the titles they want to see in more details such as weather forecasts at specific locations, up to the minute news reports, and so on. Digital terrestrial broadcasting is expected to be launched at the end of 2003 in Tokyo, Nagoya and Osaka, and throughout Japan by 2006.

These new broadcasting systems are expected to present various new services especially for hearing impaired people who want to watch captioned TV broadcasting, especially news programs. Of course, it is possible for them at present to watch captioned TV programs such as dramas through analog terrestrial TV sets with the decoders of teletext broadcasting. News programs, however, cannot be captioned because manual captioning on TV requires much time to be completed.

2. Real-time transcription system for news programs

2.1. Requirements for the real-time transcription system for news programs

All TV programs are expected to be subtitled as soon as possible. Subtitling for news programs is of critical importance, for example, important information regarding safety aspects needs to be transmitted to everyone through broadcasting. However, the speech rates are too fast for key-operators to convert correctly Japanese speech to Kana-Kanji characters in real time. Recently, it has become clear that only expert stenotypists can manage to handle the speech uttered in TV programs in real-time using special steno typewriters. The system using the steno typewriters is effective for spontaneous dialogue and for speech with environmental noises, which are difficult to be automatically recognized by just any speech recognition system.

Unfortunately, the number of such expert stenotypists in Japan is less than ten, which is not enough to broadcast daily news programs with captions. Moreover, comparing with spontaneous speech, the peculiarities of TV news speech spoken even in a quiet TV studio should be considered. They are as follows;

1. A news program includes far more location names and proper names than those in conventional speech.
2. Many varieties of topics are covered such as politics, economics, sports, culture, social affairs, foreign affairs, etc. .
3. News items which are actually covered in a news program cannot be finalized until a few minutes before the news program begins.

As expert stenotypists are not familiar with all sorts of news items, a news program will require quite a few experts. This fact implies that an alternative method of broadcasting subtitled news programs should be considered. An automatic speech recognition system is one of the candidates.

The recognition rate of the speech recognition system for speech uttered by announcers in the studio for News Programs should be no less than 95% with a recognition delay of less than 2 seconds. The recognition rate 95% is significant because the error rate 5% is the upper limit in order to correct the recognition errors in real-time by hands [Miyasaka 2001:106]

The famous speech recognition software named, as " ViaVoice" already launched on the market by IBM does not satisfy the above requirements. Due to this situation, the researchers including the author studying speech recognition system for news speech in the NHK Science and Technical Research Laboratories has decided to begin development of their own system.

2.2. Algorithms of the automatic speech recognition

Until recently, algorithms of automatic speech recognition have been globally distilled to a statistical speech recognition method including HMM (Hidden Markov Models) as the acoustic models, word bigrams (connective probabilities between adjacent two words) and word trigrams (connective probabilities between adjacent three words) as the language models. Fig.1 shows the block diagram of the system [Ando 2000:190].

Designing of the models, including an acoustic model, a language model and a decoder has been dependent upon this research.

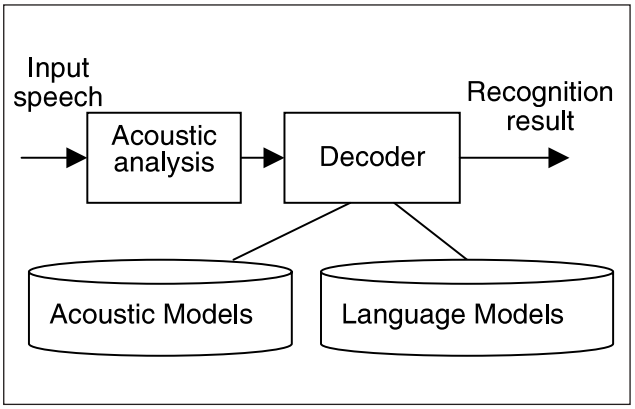

Figure 1. Block diagram of the system

2.3. The speech recognition system developed by NHK labs

The NHK study group developed by themselves a unique decoder engine (recognition algorithm), an acoustic model and a language model using the training data of up to 7 million words (1.9 million sentences) obtained by morphological analysis of the news manuscripts and transcriptions used in NHK TV news programs broadcast since 1991.

In the acoustic model as shown in Fig.2, the normal distributions of the acoustic features of each Japanese phoneme such as /a/,/i/,...., /k/,/n/...are calculated by making use of spoken data over 100 hours [Kobayashi 2001:38].

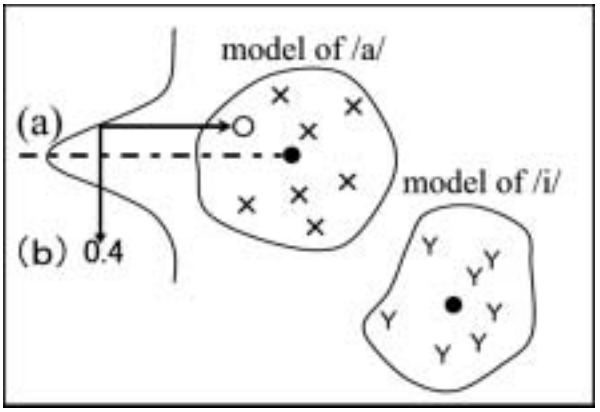

Figure 2. *Examples of some segments in the acoustic model. Xs and Ys indicate the feature vectors of phoneme /a/ and /i/, respectively. Each closed circle indicates the centroid of each model. The open circle indicates the feature vector of input data.(a) indicates the distribution of the feature vectors of the phonemes /a/. (b) indicates the probability of input data to be the phoneme /a/.*

In the language model as shown in Fig.3, the connection probability between the adjacent words is calculated by making use of the database. The probability is referred to as " word bigram". The model also uses "word trigram" to improve the accuracy.

In such statistical modeling, there is no doubt that the size and the quality of the training data for making the models will affect the performance of the system. The group has also formulated the pronunciation dictionary consisting of pairs of Kana-Kanji notations and the corresponding phoneme notations shown in Fig.4. The dictionary eliminates the manual Kana-Kanji conversion. Fig.5 shows the outline of the principle of the speech recognition system.

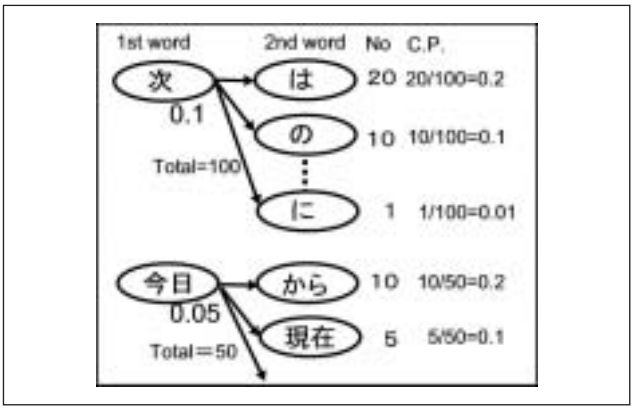

Figure 3. *Examples of some segments in the language model. "1st word" indicate the candidates of the first word of a sentence used in news. Let the number of the sentences investigated be 100 thousands and the number of the sentences* with the first word "次(Tsugi)" be 10 thousands. The probability of appearance of the word at the top of sentences

will be 0.1. The connection probability between the adjacent two words is calculated. In this figure, in succession to the 1st word "Tsugi", 100 words appear and the word "は /wa/" appears 20 times resulting in the connection probability of the word being 0.2 and so on.

Word	Pronunciation
愛	a,i
足	a,sh,i
足	s,o,k,u
…	…
次	ts,u,g,i
…	…
続いて	ts,u,z,u,i,t,e

Figure 4. An example of the Pronunciation Dictionary

The procedure is as follows;
1. Selection of the candidate of the 1st word of a sentence.
2. Preparation of the acoustic model based on the pronunciation dictionary corresponding to the selected word
3. Acoustic (Spectral) analysis of input speech waveforms
4. Extraction of the feature vector (39 dimension) per 10 msec
5. Verification and the calculation of likelihood between the feature vector and the acoustic model
6. Calculation of the acoustic score based on the verification
7. Application of the procedure above mentioned to all candidates of the words

The final candidate can be obtained based on the maximum of the acoustic score times the linguistic score (connection probabilities).

Fig.6 shows the block diagram of the decoder. The word-dependent N-best search is executed using the word bigrams as the language model in the first pass. The score of each candidate is calculated from the weighted score of the acoustic score through the acoustic models and the language score calculated using the word bigrams. In the second pass, re-scoring is performed by re-calculation of each language score using word trigrams [Imai 2000:2].

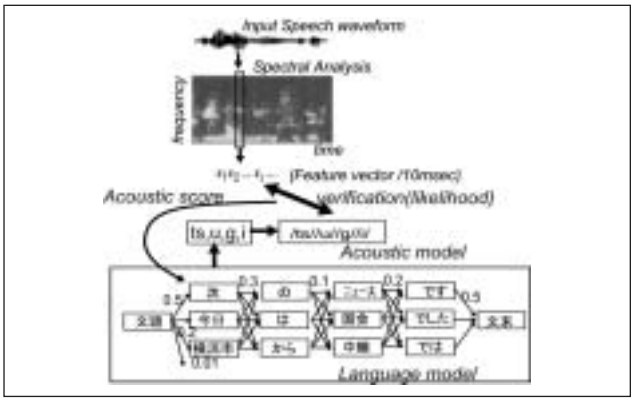

Figure 5. Outline of the principle of the speech recognition system

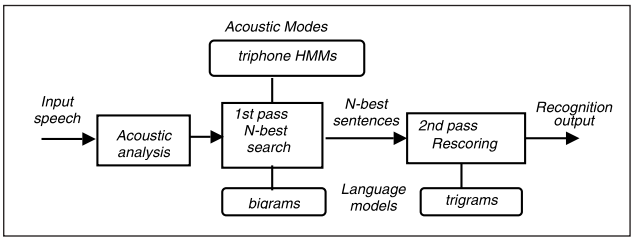

Figure 6. Block diagram of the decoder

When the connection probability is calculated with only recent news manuscripts, the probability of popular words in which appear in a week is small due to the small volume of the manuscripts. So, the probabilities of the n-gram for the words are increased using recent news manuscripts in duplicate by about 1000 times.

It will be well understood that the training data and the corresponding dictionary play important roles on constructing both the language model and the acoustic model.

2.4. Installation of the system into the news fields

One of the most important tasks imposed upon the group leader is to educate ultimate boss(es) who has full power of decision, about the quality and effectiveness of the system. The president of NHK, Mr.Ebisawa, was most impressed with the recently developed hot automatic speech recognition system which he was introduced to at the Open House of NHK's labs in May, 1999. At that time the system was continuously displaying the output characters decoded in real time from the speech uttered by an NHK male announcer sitting in a purpose-built room. The development of the system has been ongoing since then, resulting in an official announcement by the president, in October, saying that the closed-captioning service would be prepared next year on the NHK news program "News-7"

It was a surprise both for the development group and the news bureau because they had considered that it would be too fast to introduce the prototype directly into the news field and to be applied to the main and important live news program without any trials. The news staff is very sensitive in any mistakes and is also very concerned about the machine failures.

At that time, the recognition error rate was around 95% which represents the upper limit for which errors can be manually corrected in real-time. The urgent issues to be solved were as follows;
1. development of a manual error correction system
2. improvement of the reliability of the system
3. alternative methods in cases of the system failures
4. reduction of the operating costs

Fig.7 shows the real-time error correction system handled by four operators.

The correction server divides an input sentence to the terminal A or B at every pause between the adjacent sentences. The operator OP-2 or OP-4 immediately corrects the error

word pointed by the OP-1 or OP-3 who is focusing on detecting any error word based on the speech uttered by the announcer [Ando,Takagi 2001:21].

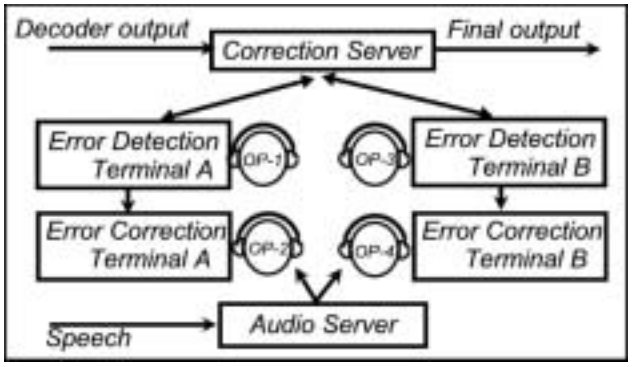

Figure 7 The real-time error correction

2.5. Development of a manual back-up system

Improvement of the reliability of the system is quite important for the development crew in order not only to operate without failure bat also to obtain the trust of the news staff who is quite sensitive to new equipment, especially when he is of the brief that it may break down at any moment.

A manual back-up system using normal key operators was also developed as an alternative to the automatic speech recognition system.

At present, news reporters prepare their preliminary drafts using PCs. Improvement of these drafts is continued until right before the corresponding news program begins. This means that the final manuscripts with which announcers use on the program are fairly different from the initial drafts. The captions made using the manual back-up system, therefore, do not much with the speech scripts uttered by the announcers because the captions are forced to be made in accordance with the drafts that are at the midterm stage. Such a back-up method, however, can reassure the nervous news crews.

The news crews are responsible for the contents dealt with in their news programs, so that correct captions without any error should be required, even though hearing impaired persons are not concerned about captions with small errors. As described before, a large number of location names and proper names appear in a news program. Even an error in those names will possess an element of danger which leads to grave infringements of human rights.

2.6. Start of the closed captioning service

At the end of March in 2000, six months after the NHK president announced to the press that the closed-captioning service would begin, the service using the hybrid system consisting of the speech recognition system and the manual system was launched on the NHK news program "News-7".

At the beginning of May, one month after the system was launched, a bus hijacking in Fukuoka District occurred in the afternoon, settling in early the next morning. The incident was broadcast live on the extended "News-7" program throughout the night. Only the automatic speech recognition system had to be used, because any preliminary manuscripts could not be prepared.

The success of live captioning by the speech recognition system earned the trust of the news crew. Since then, the manual back-up system was successfully removed. The system has been running well since then. Closed captioning making use of the improved system has been applied to sports and popular live song programs as well as news programs since 2002.

NHK's real-time closed captioning service has been regarded highly by a great number of haring impaired persons who are concerned about catching the contents of news programs in real time as well as many other programs offered by NHK.

3. Decision flow in installation of systems into the corresponding fields

3.1. Decision flow in installation of specific systems into the restricted fields

Fig.8 shows the decision flow from initial designing stage to the start up of the service introduced in a restricted corporation. The closed captioning service is one of the examples.

The flow can be described as follows;

1. An engineer or a study group proposes his idea in a closed meeting. It is necessary for him to make a rough prototype or do preliminary experiments in order to show that the idea will be effective, spending the budget allocated to the former applied plans.
2. The engineer or the group proposes his study plan in a study plan meeting in laboratories. It will be quite expected that he should explain the potential demand for the service or the system in some fields after precise future assessments as well as its technical novelty.
3. When the proposal is successful, a certain budget and members are allocated for the plan. If not, the plan should be abandoned or reconsidered.
4. The progress report should be presented at the annual review meeting. If the report is not satisfactory, the budget and the number of the study staff may be reduced or the plan itself may be abandoned.
5. The leader of the study group should make an effort to demonstrate the effectiveness of the developed prototype for his/her ultimate boss(es) as well as the director-general of the laboratories. If the demonstration is a failure, the study may be abandoned or the size may be reduced.

In the case of the speech recognition system, the demonstration was successful, so the president of NHK, who is the ultimate boss, decided that the system should be introduced in the news field after he inspected the prototype.

6. In the case of a success, the budget and the members can be increased.

In the process of development of the products or systems in NHK labs, cooperation between the labs and some manufacturers will be necessary if the labs are to be successful in finding the appropriate manufacturers in the world. In the case of the speech recognition system, the study group found that there was no such company, resulting in the development of the proprietary system.

At the demonstration, it is quite important for the study group to obtain the trust of the staff in the corresponding fields where the system may be installed. It will be common that the staff in the fields is quite sensitive to new systems which are, generally speaking, liable to break down at any moment.

The speech recognition system just developed in the laboratories was one of the typical prototypes that seem to be unstable. Fortunately, there was no objection to the plan for the system to be installed in the news field, because the staff in the fields recognized the decision of the president.

7. After the installation into the corresponding fields, various field tests will be conducted.

8. The new services will be launched.

3.2. Decision flow in installation of new systems into the common fields

Fig.9 shows the decision flow in installation of new such systems into the common fields in broadcasting as the digital HDTV broadcasting system and the digital terrestrial broadcasting system. The flow can be described as follows;

1. The items 1,2,3,4,5 and 6 described in Fig.8 are true of the case in Fig.9.

2. The basic system formats should be approved by the standard organizations such as ITU (International Telecommunication Union).

3. Formation of a consortium with manufacturers is quite important because the services require the real products such as digital HDTV receiver sets based on the approved formats. Both the talent of this formation and of the negotiation should be included in the technical strength and development capacity required by the leader of a study group.

3.3. Contribution of NHK-labs to TV broadcasting

The NHK president speaks favorably of the expression that broadcasting is a culture applying the latest technologies. There is a long history that NHK technical laboratories has served new broadcast technologies.

Table.1 shows the main contribution of NHK labs to TV broadcasting since 1945. The table indicates that epoch making new broadcasting services have been based on the broadcasting systems developed in the labs.

The digital HDTV broadcasting, for example, which has been launched for the first time in the world in December 2000, originated in HDTV born in the labs. There is a history that the HDTV technology was rejected from both Europe and America fearing the domination of the world market for television in 1980's although HDTV visual images with high quality attracted viewers in the worldwide exhibits. It took 30 years that the 1125 scanning line HDTV studio format proposed by NHK was approved as the Unification studio format by ITU.

3.4. Some important issues in designing

There can be several reasons that users in the fields may often be conservative in the introduction of new systems and/or equipment into the fields. These are as follows;

1. The users have become familiar to operate the existing systems and/or equipment, so that they want to improve their expertise, but they often do not want to change the circumstance.

2. Once the system is introduced into the fields, the users will be afraid to take the risks because they have to be responsible for operating the system without any mistake or error even if the system has any bugs.

3. They may fear losing their jobs after the system will be introduced into their fields.

It will be important that developers should consider the users' potential reservations mentioned above in designing the systems concerned.

4. Conclusion

There are some problems that may occur in the process of introduction of new systems and/or equipment just developed into broadcasting fields. This paper presents a case in the introduction of the automatic speech recognition system developed in technical laboratories into news fields as one of the examples.

In the decision flow from initial designing stage to the start up of the services in broadcasting fields, various problems to be solved are considered. One of the most important issues to be settled is that the developers should obtain the trust of the staff in the fields where a new system will be installed.

Any broadcasting new service inevitably affects public viewers although it will take a long time until the service will be widely accepted. In the case of the satellite broadcasting service in Japan, for example, it took more than ten years before ten million viewers signed on.

The reason why it will take a long time to demonstrate the success of the service is that the value exists not only in the excellent technical systems such as HDTV receiver sets but also in the amount of high-quality programs suited for the systems.

On the contrary, consumers, for example, can quickly evaluate the value of personal computers. In such cases, So called "User-centered design (UCD)" presented by IBM will be one of excellent methodologies for designing [IBM :see URL]. It can be said that the value of a new system or equipment

depends on the specific features of the system to be developed. One of the problems to be considered in the future revolves around what is the most effective methodology for introducing new systems successfully into the corresponding fields.

References

1)Miyasaka,E, Imai,A, Takagi, Imai,T and Andi,A, Application trials for the realization of human-friendly broadcasting, J. Gerontechnology, 103-110(2001).

2)Ando,A, Imai,T, Kobayashi,A, Isono,H and Nakabayashi,K, Real-time transcription system for simultaneous subtitling of Japanese broadcast news programs, IEEE Trans. Broadcasting, Vol.46,189-196.(2000).

3)Kobayashi,A, Speech recognition system for closed captioning of broadcast news, Proceedings of Public Open House of NHK Science and Technical Laboratories, 37-42 (2001)

4)Imai,T., Kobayashi,A., Sato,S., Tanaka,H., and Ando,A, Progressive 2-pass decoder for real-time broadcast news captioning, Proceedings of International Conference on Acoustics, Speech and Signal Processing, 1937-1940(2000)

5)Ando,A and Takagi,T, Automatic closed captioning system for broadcast news speech, NHK STRL R&D, No.65, 16-21(2001)

6)http://www-106.ibm.com/developerworks/usability/library/us-ucd/

1950	First open experiment monochrome TV broadcasting after World War II
1953	Start up of study on color TV broadcasting
1956	Open experiment of color TV broadcasting
1960	Start up of color TV broadcasting
1964	Success of transmission of Tokyo Olympic Games via satellite
1965	Start up of study on High Definition TV (HDTV)
1966	Start up of study on Satellite broadcasting
1970	Exhibition of first prototype of HDTV
1971	Start up of study on wall-mounted flat plasma display panel (PDP)
1978	Test experiment of sound multiplex broadcasting
1982	Development satellite broadcasting system
1982	Start up of study on Integrated Service Digital Broadcasting (ISDB)
1984	Development of analogue transmission system for HDTV (MUSE system)
1984	Test experiment of satellite broadcasting
1986	Start up of experiment of satellite HDTV broadcasting
1989	Start up of satellite broadcasting
1989	Start up of experimental satellite HDTV broadcasting
1991	Start up of test satellite HDTV broadcasting
1994	Start up of practical satellite HDTV broadcasting
1996	Start up of experiment of ISDB
1998	Development of BS digital HDTV broadcasting system (BS: Broadcasting Satellite)
1999	Development of Terrestrial digital HDTV broadcasting system
1999	Development of real-time News Transcription system by the automatic speech recognition system
2000	Start up of closed captioning service at News-7 using the real-time News Transcription system
2000	Approval of unification HDTV studio standard presented by NHK by the ITU (International Telecommunication Union)
2002	Development of 4000 scanning line ultra high definition image system

Table1 Main contribution of NHK Science & Technical Labs to TV broadcasting