

A Report on a Stay as Visiting Student at Tokyo Institute of Technology

Matthias Hietland Heie

Norwegian University of Science and Technology

Acknowledgements

This report is based on my stay as visiting student at Tokyo Institute of Technology from April to September 2003. There are many who deserve my gratitude from making this stay an amazing experience. I would like to thank Japanese Student Services Organization for the generous scholarship I was awarded. I am deeply grateful to Professor Furui for having accepted me as a visiting student at his laboratory. Studying this interesting topic under his supervision has been an invaluable experience. Without the extraordinary help of Ms Dale Licata at International Office, NTNU, and her counterpart at TIT I would never have succeeded in my attempt to get accepted as a visiting student in Japan. Thank you very much. Last, but not least, I would like to thank all members of Furui Laboratory, TIT, for help, inspiration, encouragement and friendship.

Table of Content

Acknowledgements	2
The Visit	4
1.1 Reasons for Studying in Japan	4
1.2 The Application Procedure	5
1.3 The Language	5
1.4 Academic Facilities and Environment.....	6
1.5 Administrative Issues	7
1.6 Housing	7
1.7 Economy	8
1.8 Cultural and Social Life in Tokyo	8
1.9 Consequences and Conclusion	8
The Research	10
2.1 The System	10
2.2 Scoring System.....	11
2.3 Sentence Extraction	14
2.4 Sentence Compaction	14
2.5 Experimental Setup	15
2.6 Results and Discussion	16
2.7 Conclusion.....	18
Bibliography	20

Part 1

The Visit

In this part I will describe the administrative, academic and social circumstances prior and during my visit to Tokyo Institute of Technology (TIT) from April to September 2003. During this period I conducted research for my sivilingeniør (MSc) thesis. I was placed at Furui Laboratory, under the supervision of Professor Furui. The area of my study was automatic speech summarization.

I started the 5 year sivilingeniør program at NTNU in 1998, and after three years of general computer science studies I began specializing in computational linguistics, speech processing and other artificial intelligence subjects. The fourth year of my study (2001/02) I spent at Delft Technical University (TU Delft). Thus my stay at TIT was my second time being a visiting student at a foreign university.

1.1 Reasons for Studying in Japan

Between my compulsory army service year and university, I had a gap year where I travelled 6 months in South East Asia. During this trip I developed a strong interest in Asian culture. So after my third year in university, when I decided to apply for an internship abroad, it was natural for me to apply for positions in Asia. The international student organization IAESTE organizes internship exchanges for students at technical universities, and I applied for several of their offered internships, all of them in East Asia. I was successfully chosen for a position at NTT R&D in Yokosuka, Japan, living at NTT's dormitory in Yokohama. I spent 6 weeks at their image processing laboratory. The stay was an amazing experience for me. I enjoyed very much working with my Japanese colleagues, I made good friends among the Japanese IAESTE members (students from various universities in Tokyo) and I found the Tokyo-Yokohama area an incredibly interesting place. The advantages of this stay were two-folded:

it boosted my motivation for computer science and it enhanced my interest in East Asian culture, from now on especially in Japanese culture.

During my exchange studies at TU Delft, I investigated the possibilities for conducting the research for my Master's thesis in Japan. The professor at the laboratory where I studied suggested that I should contact a Japanese professor he knew at TIT, Professor Furui. I sent him a letter, and he was kind enough to accept me at his laboratory. Now that I had an invitation from a Japanese professor, I could start the procedure of applying for a stay as visiting student in Japan.

1.2 The Application Procedure

When I came back to NTNU after my stay at TU Delft, I started the application procedure. First I contacted Ms Dale Licata at NTNU's international office, who provided me with all the necessary information in order to file an application. First of all I needed a professor at NTNU to be my home supervisor. Such an agreement is a pre-requisite for conducting the research for the thesis outside the university. Professor Tore Amble at my department (Department of Computer Science) was willing to be my home supervisor. There was a considerable amount of administrative work to be done in order to file the application. Fortunately for me I was spared most of the work due to the fantastic job done by Ms Licata and her Japanese counterpart. I applied for both a place at TIT, and a scholarship from JASSO. To my great relief and gratitude, the answers were positive, I was approved both a place at TIT and the scholarship.

I finished my coursework in December 2002, but could not attend TIT before the beginning of the spring semester, April 2003. Therefore I wanted to start a language course as soon as possible. But it took quite a while to receive all the necessary immigration papers in order to get a Japanese student visa. As soon as I received these papers, the Japanese embassy in Norway was extremely helpful, and speeded up the visa process to issue me a visa as quickly as possible. I finally entered Japan in early February.

1.3 The Language

Languages have always interested me, so when it was time to specialize inside computer science, I combined my computer studies with my language interest by choosing computational linguistics. And when I got the chance to study in Japan for 6 months, it was natural for me to seize the opportunity to learn some of the language. Japanese being unrelated to any of the languages I already knew, I understood that it would be limited how

much I could expect to learn. But because I had nothing else to do between January and April, I decided to spend as much time as possible studying Japanese prior to my studies. When I finally was ready to enter Japan in early February, I had booked 6 weeks at Yamaza School of Japanese, a non-profit school in Okazaki, a city near Nagoya. The classes were purely focused on conversation, which suited me well. When I started, there were no other students at entry level, so I had one teacher only for me. This proved to be an extremely efficient way of learning, and by the end of the course, I was able to have basic conversations in Japanese. Being far from fluent in Japanese was not a major obstacle during my stay in Japan. The research at my laboratory was conducted in English, and most members of the laboratory were quite good in English. Outside university it was not always easy to get by, but normally my limited knowledge of Japanese was enough to sort out most issues. Still, I can easily imagine that a foreigner who spends longer time in Japan than I did, will start to feel isolated if (s)he doesn't acquire more advanced Japanese skills.

1.4 Academic Facilities and Environment

The academic facilities at TIT were excellent. At my laboratory I had my own modern computer with LCD screen. There were a lot of relevant academic literature there, and even more at the library. The academic environment was professional and inclusive, and the relationship with the academic staff at the laboratory was very good. Professor Furui is highly regarded in the field of speech processing, and I felt honoured to have him as my supervisor. Whenever I had an academic problem, I was free to consult with either him or the assistant professor at the laboratory, Professor Iwano. Each week we had a laboratory meeting, where everyone presented the status of their research.

What I particularly enjoyed was our daily lunches, where the whole laboratory (or those who chose to) went to the canteen to eat lunch together. This included professors, post doc researchers, PhD students, MSc students, last year undergraduate students and administrative staff. At my laboratory at TU Delft, we had the same tradition, but at NTNU I have never had lunch with a professor. This might seem like a small issue, but I really appreciated the opportunity to meet with the academic staff under such informal circumstances and it made me feel part of a professional academic environment.

There were courses offered in English at TIT, but I did not attend any of them since the purpose of my stay was research.

1.5 Administrative Issues

The two main administrative issues for me were housing and my JASSO scholarship. Both of these issues were solved smoothly with the help from the international office at TIT. In order to receive the scholarship, I had to open a bank account, which I did with the help of my laboratory mentor. Further, I had to sign every month at the international office to prove that I was present. My laboratory had its own secretary, and whenever there was an administrative problem, she was very helpful.

Two other issues were to register as a resident of Tokyo and register for national insurance (which was inexpensive for foreign students). One morning all new residents of my dormitory went to the municipal office where both these tasks were taken care of, which was a relief to us as newcomers to Tokyo.

All in all I would say that there were never any major administrative problems during my stay at TIT, much thanks to the professionalism of the staff of TIT.

1.6 Housing

During my entire stay in Tokyo I stayed in one of the dormitories for international students. The dormitory was situated in Komaba-Todaimae, near Tokyo University's campus. This was a central location, two train stops (or 20 minutes walking) from Shibuya. Still the area had a sub-urban feeling. The price was very low for such a central location. I am certain it was not market price, but subsidised for foreign students. The area consisted of several dormitory buildings, each with around 40 rooms. Each room (all single rooms) had its own bathroom, and there was one kitchen per floor (about 15 students per kitchen). Each building also had a TV room with daily newspapers (in both Japanese and English). The dormitories were modern, in good condition and with facilities such as washing and drying machines, and broadband internet connection. This being a dormitory for international students, all students were foreigners, except for about 5 Japanese mentors, who arranged various events for us. I was very pleased with the housing offered. It had all the facilities I needed, and it was easy to get contact with other students in the communal areas. However, it would have been nice with some more Japanese students around. Another issue was the midnight curfew. Being 25 years old and having lived on my own for the last 6 years, it was a bit odd to have to be home by a specific time.

1.7 Economy

Tokyo is ranked as the most expensive city in the world. Fortunately the Japanese government is one of the most generous in the world when it comes to awarding scholarships to foreign students. I had been awarded a scholarship from JASSO, which meant that every month I would receive 80 000 yen. If this had been my only source of financing, my budget would have been very tight. But I had other sources of financing, and as mentioned before, my dormitory rent was subsidised. Coming from Norway I am already used to a high price level, so in total I found myself having more money to spend than I have ever had before. This gave me the opportunity to travel around in the weekends and enjoy Japanese foods in the restaurants.

1.8 Cultural and Social Life in Tokyo

Tokyo is one of the most interesting cities I have been in, and to actually live here was an amazing experience for me. The daily life in Tokyo was convenient, with supermarket nearby my dormitory and reliable transport to the campus. The trains were a bit more cramped than I was used to, only natural in a city of such a size. I cannot recall ever having experienced a delay. The only problem was in the nights. All transport closed down at midnight, except for taxis, which were very expensive. This made it a bit difficult to go out in the weekends. Fortunately my dormitory was in walking distance from Shibuya, which has a lot of nightlife to offer.

I enjoyed spending the weekends investigating the different areas of the city. Each area had its own character. The cultural scene in Tokyo was brilliant. There were lots of cinemas, theatres, galleries, concert ventures etc.

Before I started my studies at TIT I already had friends in Tokyo: students from IAESTE whom I met during my internship one and a half year earlier. These people, and international students that I met at the dormitory, became my best friends during my stay. I did not have much contact with other TIT students during my stay. An important reason for this was that I focused more on my studies than ever before. I felt that my thesis was the last lap of a 5 year run, so I put most of my effort into the research, rather than having a hectic social life.

1.9 Consequences and Conclusion

The 6 month study visit at TIT was a great experience I wouldn't be without. My research project was very interesting, and my motivation was high. I was extremely satisfied with the

academic surroundings, and I could feed my interest for Japanese culture. To others who might consider visiting Japan for study purposes, I will definitely recommend it. They should even consider studying to a whole degree there, for example doing a MSc. The Japanese government has very generous scholarship for foreigners who succeed to qualify for such studies.

The only disadvantage I can think of was being away from the Norwegian recruitment market. Many of my fellow students in Norway used the last semester of their study to apply for jobs. Being in a foreign country, this was very hard for me. I did a few attempts, but found it too difficult. This was not entirely negative, instead of thinking about job hunting I could put all my efforts into my research. But when I came back to Norway, I had no contacts in the labour market. During my stay in Tokyo I decided to try to find a job in London, so shortly after my arrival home, I moved to London. At this time it was very difficult to find work in the IT sector, especially without work experience. But after a few months I found work, and now I am working for a small software house in London.

It is my intention to go back to Japan, either to do further studies or to find work. I know the latter is difficult without good Japanese skills, but I am currently using my spare time to study Japanese.

Part 2

The Research

This part describes the research I conducted during my stay at TIT. Based on this research I wrote a thesis[1] which was submitted to NTNU in October 2003 for the degree of sivilingeniør (MSc).

In previous work [2, 3] a method for speech summarization of English broadcast news, based on word extraction, was proposed, and its effectiveness verified. This work is an attempt to increase the quality of the summaries by solving one of the main problems with the existing approach: the generation of small and incomplete sentences.

2.1 The System

In the existing system, from each sentence a set of relatively important words are extracted and concatenated to form a summary of the source sentence. The size of the extracted set is given by a user-specified summarization ratio. The summarized sentences are then concatenated to form a summary of the source article. What words to extract are decided based on a sentence compaction score, which weights the possible sets according to their information value plus linguistic and acoustic features.

This method was further extended by letting the summarization ratio for each sentence be dynamically decided, while the summarization ratio for the source article as a whole is decided by the user-given summarization ratio. Multiple sentences are summarized by finding the set of sentence-level summaries that maximizes the sentence compaction score, within the bounds of the summarization ratio. The motivation for this extension is the fact that not every sentence is equally important, and by letting the sentence-level summarization ratio vary dynamically, the most important sentences will be given preference to, on expense of less important ones.

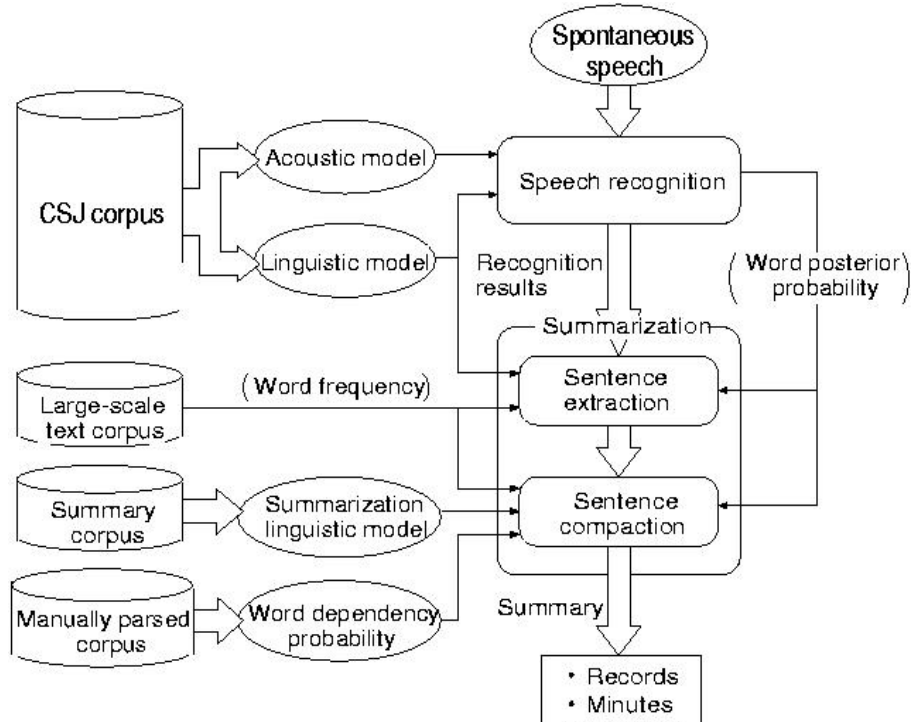


Fig. 1 System architecture [4]

The problem with this method, especially when the summarization ratio is small, is that the summaries tend to include very short sentences that are incomplete and therefore difficult to read.

To deal with this problem an extension of the existing method was proposed for Japanese [4, 5]. The summarization now consists of two stages, which are illustrated in figure 1:

1. *Sentence extraction stage*: the most relatively important sentences are extracted based on a sentence extraction score.
2. *Sentence compaction phase*: words are extracted from the extracted sentences in the same way as in the existing system.

This method was tested for Japanese presentations and its effectiveness confirmed. In this work the same method will be tested for English broadcast news.

2.2 Scoring System

What sentences to extract in the sentence extraction stage, is decided by the sentence extraction score. Similarly, what set of words to extract from each sentences in the sentence compaction stage is decided by the sentence compaction score. These scores are sums of weighted subscores, and although these two scores are not calculated equally, they still share some subscores. These subscores will be described below, while how these scores are used to

calculate sentence extraction score and sentence compaction score will be explained in the following section.

Word Significance Score

The word significance score I for a topic word indicates its significance, i.e. the relative information value of this word in the transcript. The purpose of this measure is to include the most important topic words in the summary, while disregarding less important ones, and is hence used in both stages. The word significance score is given by the term frequency multiplied with the inverse document frequency ($tf * idf$):

$$I(w_i) = f_i \log \frac{F_A}{F_i}$$

where:

- w_i : a word in the transcribed speech
- f_i : number of occurrences of w_i in the transcribed article
- F_i : number of occurrences of w_i in the training data
- F_a : number of occurrences of all topic words in the training data (i.e. $\sum_i F_i$)

Nouns and verbs are chosen as topic words. Other words are awarded a flat score. Each reappearing topic word is also given a flat score to reduce repetition of words in the summarized sentence. This approach rewards words that are relatively more frequent in the source article than in the training data, as these words are thought to be more significant.

Linguistic Score

The linguistic score L indicates the linguistic likelihood of a word string: $P(w_1, w_2, \dots, w_M)$. As seen in chapter 2, this can be approximated by using n-grams, in this case trigrams:

$$P(</s> | w_M, w_{M-1}) * P(w_M | w_{M-1}, w_{M-2}) * P(w_{M-1} | w_{M-2}, w_{M-3}) * \dots * P(w_1 | <s>)$$

where $<s>$ is the start symbol and $</s>$ is the end symbol.

By including this score in the sentence compaction score, more likely sentences will be rewarded. This increases the readability and grammatical correctness of the summarized sentences. In sentence extraction full sentences are extracted. Still this score is included because sentences with a higher linguistic probability have a lower probability of containing recognition errors.

Confidence Score

Since speech recognition is not perfect, a transcript will contain recognition errors. These errors should not be included in a summary. To account for recognition errors, a confidence

score C for each word is calculated. This score is the posterior probability of the given word, including both acoustic and linguistic likelihood. The posterior probability of a word is the ratio of the word hypothesis probability to the probability of all hypotheses. These hypotheses are obtained by the decoder in the ASR component, and can be represented by a word graph. An example of such a word graph is shown in figure 2. In reality word graphs tend to have a lot more hypotheses, and are therefore more complex than in this example. S and T are the beginning node and the end node. Every other node represents points in time, while the links connecting these nodes represent word hypotheses. Every link is associated with the acoustic likelihood and the linguistic likelihood of the word hypothesis.

The posterior probability of a word hypothesis $w_{k,l}$ is given by:

$$C(w_{k,l}) = \log \frac{\alpha_k P_{ac}(w_{k,l}) P_{lg}(w_{k,l}) \beta_l}{\zeta}$$

where:

- k, l node number in a word graph
- $w_{k,l}$ word hypothesis between node k and node l
- $C(w_{k,l})$ log of the posterior probability of $w_{k,l}$
- α_k forward probability from the beginning node S to node k
- β_l backward probability from node l to the end node T
- $P_{ac}(w_{k,l})$ acoustic likelihood of $w_{k,l}$
- $P_{lg}(w_{k,l})$ linguistic likelihood of $w_{k,l}$
- ζ forward probability from the beginning node S to the end node T

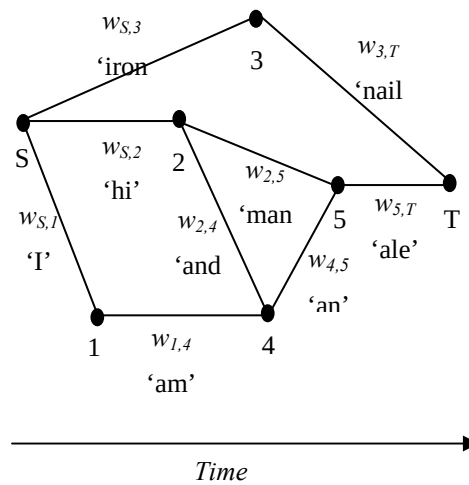


Fig.2 An example of a word graph

Word Concatenation Score

The concatenation score is included to ensure that the semantics of the original sentence is maintained, and is only used in the existing compaction part of the system. This score weights a word concatenation between words with dependency in the transcribed sentences. The dependency is measured by a phrase structure grammar, SDCFG (Stochastic Dependency Context Free Grammar) [2].

2.3 Sentence Extraction

The most relevant, well-structured and correctly recognized sentences are extracted from the text. The sentence extraction ratio (*SER*) yields the ratio of the transcription that is to be extracted in this first stage.

Extraction of sentences is done according to the sentence extraction score. This score is calculated for every sentence $W = w_1, w_2 \dots w_N$ where w_k is the k -th word in the sentence and N is the sentence length:

$$S(W) = \frac{1}{N} \sum_{i=1}^N \{L(w_i) + \lambda_I I(w_i) + \lambda_C C(w_i)\}$$

where $L(w_i)$, $I(w_i)$ and $C(w_i)$ is the linguistic score, significance score and confidence score, respectively. λ_I and λ_C are weighting factors. It should be noted that the purpose of the linguistic score in the sentence extraction step is different from the purpose of the linguistic score in the sentence compaction step. In the sentence compaction step linguistic score is included in the calculation to generate sentences that are as readable as possible. In sentence extraction, however, the sentences are extracted as a whole for further processing by sentence compaction, hence there is no sentence generation. Still the linguistic score is included also in sentence extraction, to de-weight linguistically unnatural word strings caused by recognition errors. Word concatenation score is not included in the sentence extraction score. In sentence compaction this score is included to generate sentences that are as semantically equal to the original source sentence as possible. Since every extracted sentence is extracted in its entirety, the meaning of the sentence will be preserved. Hence this score is obsolete.

2.4 Sentence Compaction

To summarize a sentence for a given compression ratio, an ordered subset of words from the source sentence that is smaller than the ratio and that maximizes a summarization score, is extracted. This ordered set is concatenated to constitute the summary of the source sentence. The resulting summarized sentence should keep important information while less important

information is deleted. Further, recognition errors should not be present, while the sentence should be as readable as possible and the meaning of the summarized sentence should not be inconsistent with the source sentence.

Given a source sentence with N words, $W = w_1, w_2, \dots, w_N$, summarization is performed by extracting a set of M words, $V = v_1, v_2, \dots, v_M$, that maximizes the sentence compaction score. The sentence compaction score of each candidate is calculated as follows:

$$S(V) = \sum_{m=1}^M \{I(v_m) + \lambda_L L(v_m | v_{m-1}, v_{m-2}) + \lambda_C C(v_m) + \lambda_T T(v_{m-1}, v_m)\}$$

where L , C and T is the linguistic score, confidence score and word concatenation score respectively, as described above. λ_L , λ_C and λ_T are weighting factors.

An algorithm based on dynamic programming is used to find the candidate summarization that maximizes this score.

Each sentence is not equally important. The amount of information varies between sentences. In addition some sentences contain more recognition error than others. Therefore summarizing each sentence individually for the target compression ratio does most likely not result in the optimal summarization result.

To handle this problem a two-level dynamic programming technique is used. This technique summarizes each sentence for every possible number of target words, using the algorithm from last section multiple times per sentence. The next step is to find the sequence of individual sentence summaries that does not exceed the target compression ratio and that maximizes the summarization score.

2.5 Experimental Setup

The same experimental setup as in [3] was used.

Evaluation Data

Five TV news broadcasts in English from CNN recorded in 1996 were tagged with Brill's tagger [6]. These broadcasts consisted of 25 utterances on average. Seventeen native English speakers generated manual summaries by extracting words from manual transcriptions of the broadcasts. These summaries were merged into a word network. Two such word networks were built: one with a summarization ratio of 40% and one with 70% ratio.

Speech Recognition System

The news broadcasts were transcribed by the JANUS speech recognition system [7]. Speech was digitized at 16 kHz sampling and 16 bit quantization. Feature vectors had 13 elements consisting of MFCC. VTLN and cluster-based cepstral mean normalization were used to compensate for speakers and channels. LDA was applied to produce a 42-dimensional vector from a set of features in each segment consisting of 7 frames. The acoustic model was a pentphone model with 6000 distributions sharing 2000 codebooks. There were about 105 k gaussians in the system. The training data was comprised of 66 hours of broadcast news. The language model consisted of bigram and trigram models, and was based on a broadcast news corpus with a vocabulary of 40 k. A word graph based 3-pass decoder was used for transcription

Training Data for Summarization Models

A word significance model and a bigram language model was created using 175431 sentences (about 4.7 M words) from the Wall Street Journal corpus and the Brown corpus in the Penn Treebank [8]. The SDCFG was constructed using 10681 sentences from the Brown corpus.

Experiments

The transcriptions were summarized for two different summarization ratios: 40% and 70%. For each ratio, the transcriptions were summarized for eleven different sentence removal ratios (*SRR*): 0, 10, 20, ... ,100. As explained in chapter 3 there are 3 subscores used in the sentence extraction score: linguistic score (*L*), word significance score (*I*) and confidence score (*C*). Thus the process was repeated for each combination of sentence extraction subscores, i.e. *L_I_C*, *L_I*, *L_I*, *L_C*, *I_C*, *L*, *I* and *C*. In addition the experiments was conducted for randomly extracted sentences (*Rand*). The evaluation was executed using leave-one-out cross-validation [9], in which the set of articles were split into training sets consisting of four articles and test set with one article. Hence there were five combinations of training set/test set pairs. Weights of the sentence extraction subscores were optimized on the training sets and applied on the test sets. Weights of the sentence compaction subscores were held constant at optimal level found at *SRR* = 0 during the whole experiment.

2.6 Results and Discussion

The results of the experiments are depicted in figure 3 and figure 4. It can clearly be seen that the two-stage speech summarization method is better than the previous summarization method consisting solely of sentence compaction (*SRR* = 0). In the case of 40% summarization ratio the summarization accuracy score rises from 54 with only sentence

compaction to 75 for the best combination of subscores and sentence removal ratio. With a summarization ratio of 70%, the increase is smaller but still significant: from 71 to 81. In the 40%-case the best results are achieved by a sentence removal ratio of 100. This means that the two-stage summarization method is reduced to a one-stage method, consisting only of sentence extraction and no sentence compaction. In the 70%-case the best conditions are achieved with a sentence removal ratio of about 80-90. The sentence extraction stage still clearly dominates the sentence compaction stage, but the optimum is reached by combining the two methods into the proposed two-stage method. That the two summarization ratios differs in this matter, was expected. Extracting words from many sentences will lead to a high number of small sentences if the summarization ratio is small. As earlier explained, these small sentences tend to be unnatural and incomplete, and will therefore give a low summarization score. But that the extraction phase would be so dominant, also for higher summarization accuracies, came as a surprise. It is worth noting that in the 40%-case extracting sentences randomly increases the accuracy, a phenomenon that is not present in the 70%-case.

An explanation of this domination by sentence extraction might be that humans summarize mainly by picking out the most salient sentences, i.e. sentence extraction. But an additional reason could be that in most cases there will be a path through the word networks going through every (or most) words, and therefore full sentences will tend to be favoured by the summarization accuracy method.

If one consider the different set of subscores used in the summarization process, the summaries can be grouped into three clusters for both summarization ratios:

1. Confidence score included (L_I_C , L_C , I_C and C)
2. Confidence score excluded, significance score included (L_I and I)
3. Only linguistic score included, or sentences randomly extracted (L and $Rand$)

The confidence score appears to be the single most important extraction subscore. Including significance score gives better performance than extracting sentences randomly, so this subscore has a positive effect on its own. However, no improvement is achieved by combining this score with the confidence score. The linguistic score has a small positive effect on its own, but no significant improvement is achieved by including this subscore with the other subscores. This was not unexpected. The motivation for including this score was to de-weight unnatural sentences in the transcription caused by recognition errors. But this is already handled by the confidence score. However one could have expected a significant improvement by expanding I to L_I . But this is not achieved. It was also unexpected that the combination of the confidence score and the significance score did not yield better results

than using only confidence score. That the confidence score is so dominant can be explained by the importance of leaving out recognition errors. But an additional cause might also be that the summarization accuracy method is very sensitive to recognition errors.

2.7 Conclusion

In this method, inadequate sentences including recognition errors and less important information are automatically removed before wordbased sentence compaction. It was confirmed that in spontaneous presentation speech summarization, combining sentence extraction with sentence compaction is effective, and the new method achieves better summarization performance than our previous one-stage method. It has also been found that the word significance score and the word confidence score are effective for extracting important sentences. The two-stage method is effective for avoiding the production of short unreadable sentences, one of the problems that the one-stage method had.

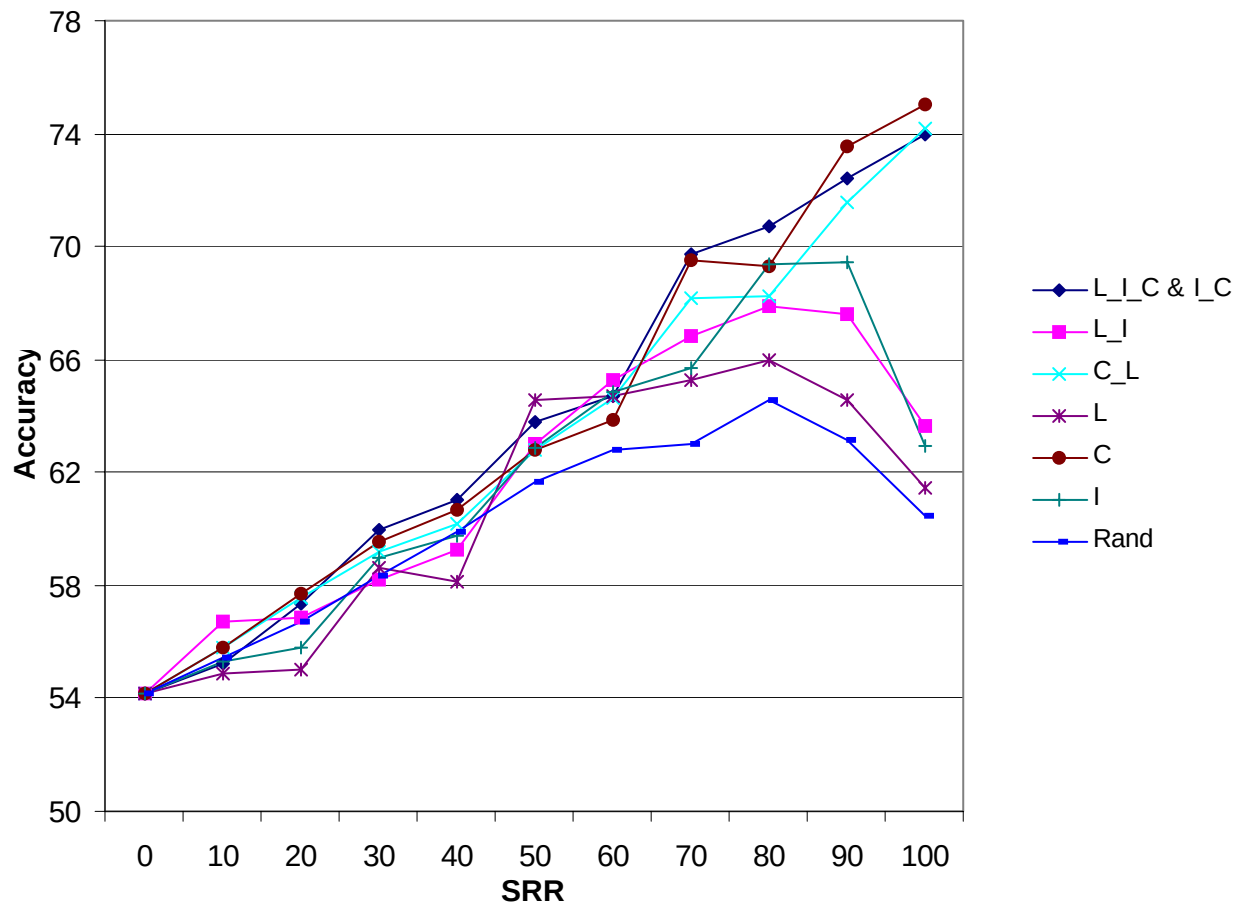


Fig. 3 Summarization accuracy for 40 % summarization ratio

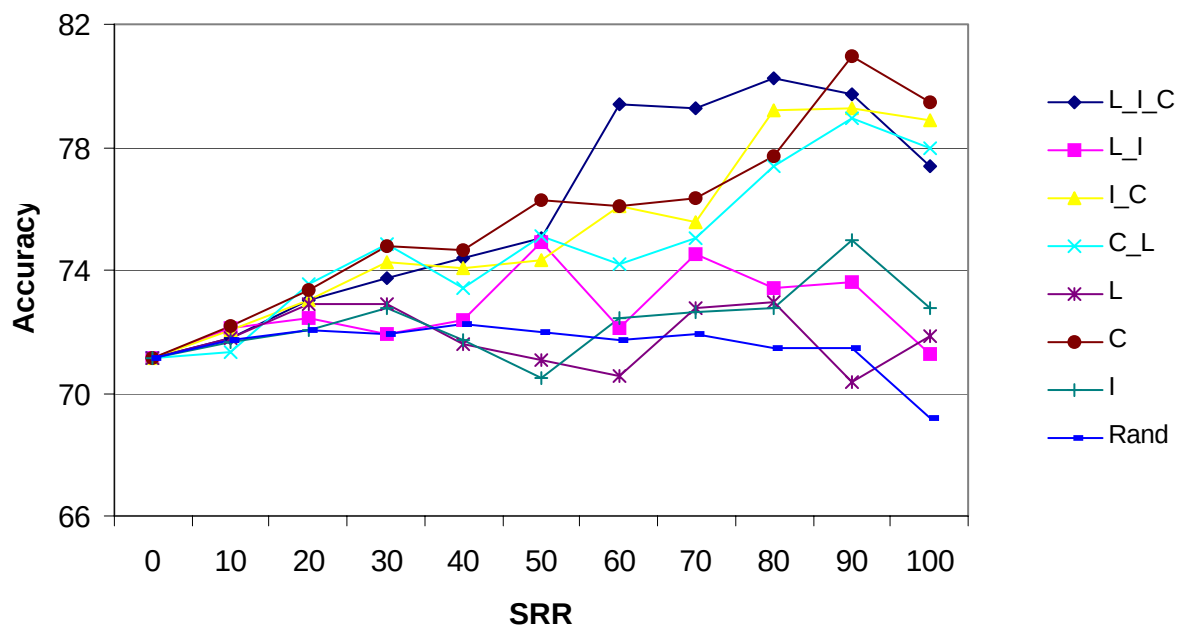


Fig. 4 Summarization accuracy for 70 % summarization ratio

Bibliography

- [1] M. Heie, “A Study on Automatic Speech Summarization for English using Sentence Extraction and Sentence Compaction”, *Sivilingeniør thesis submitted to Norwegian University of Science and Technology*, 2003
- [2] C. Hori, “A study on Statistical Methods for Automatic Speech Summarization”, *Doctoral thesis submitted to Tokyo Institute of Technology*, 2002
- [3] C. Hori, S. Furui, R. Malkin, H. Yu and A. Waibel, “A Statistical Approach for Automatic Speech Summarization”, *EURASIP Journal on Applied Signal Processing*, 2003:2, p. 128-139, 2003
- [4] T. Kikuchi, S. Furui and C. Hori, “Two-Stage Automatic Speech Summarization by Sentence Extraction and Compaction”, *Proc. of ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition (SSPR)*, pp. 207-210, 2003-4
- [5] T. Kikuchi, S. Furui and C. Hori, “Automatic Speech Summarization Based on Sentence Extraction and Compaction”, *Proc. of IEEE International Conference on Acoustics, Speech, Signal Processing*, vol. 1, pp 384-387, 2003
- [6] <http://www.cs.jhu.edu/~brill>
- [7] A. Waibel et al., “Advances in Meeting Recognition”, *Proc. of the 1st International Conference on Human Language Technology Conference*, pp. 11-13, 2001
- [8] <http://www.cis.upenn.edu/~treebank>
- [9] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*, Wiley, 1972