Title	Automatic Summarization of English Broadcast News Speech
Author	Chiori Hori, Sadaoki Furui, Rob Malkin, Hua Yu, Alex Waibel
Journal/Book name	Human Language Technology 2002, Vol. , , pp. 241-246
発行日 / Issue date	2002, 3

Automatic Summarization of English Broadcast News Speech

Chiori Hori†, Sadaoki Furui†, Rob Malkin‡, Hua Yu‡ and Alex Waibel‡

†Department of Computer Science, Tokyo Institute of Technology,
2-12-1, O-okayama, Meguro-ku, Tokyo, 152-8552 Japan
{chiori,furui}@furui.cs.titech.ac.jp
‡Interactive Systems Labs, Carnegie Mellon University, Pittsburgh, PA 15213, USA
{malkin,hua,ahw}@cs.cmu.edu

ABSTRACT

This paper proposes an automatic speech summarization technique for English. In our proposed method, a set of words maximizing a summarization score indicating appropriateness of summarization is extracted from automatically transcribed speech and concatenated to create a summary. The extraction process is performed using a Dynamic Programming (DP) technique according to a target compression ratio. In this paper, English broadcast news speech transcribed using a speech recognizer is automatically summarized. In order to apply our method, originally proposed for Japanese, to English, the model of estimating word concatenation probabilities based on a dependency structure in the original speech given by a Stochastic Dependency Context Free Grammar (SDCFG) is modified. A summarization method for multiple utterances using two-level DP technique is also proposed. The automatically summarized sentences are evaluated by a summarization accuracy based on the comparison with the manual summarization of correctly transcribed speech by human subjects. Experimental results show that our proposed method effectively extracts relatively important information and remove redundant and irrelevant information from English news speech.

Keywords

Speech summarization, Summarization scores, Two-level Dynamic Programming, Stochastic Dependency Context Free Grammar, Summarization accuracy

1. INTRODUCTION

Recently, large-vocabulary continuous-speech recognition (LVCSR) technology has made significant advancement. Real time systems can now achieve word accuracy of 90 % and above for speech dictated from newspapers. Currently various applications of LVCSR systems, such as automatic closed captioning [1], meeting/conference summarization [2][3] and indexing for information retrieval [4], are actively investigated. Transcribed speech usually includes not only redundant information such as disfluencies, filled pauses, repetitions, repairs, and word fragments, but also irrelevant information caused by recognition errors. Therefore, practical applications using LVCSR systems require a process of speech summarization which removes redundant and irrelevant information and extracts relatively important information depending on users' requirements, especially for spontaneous speech.

HLT '02 San Diego, California USA
.

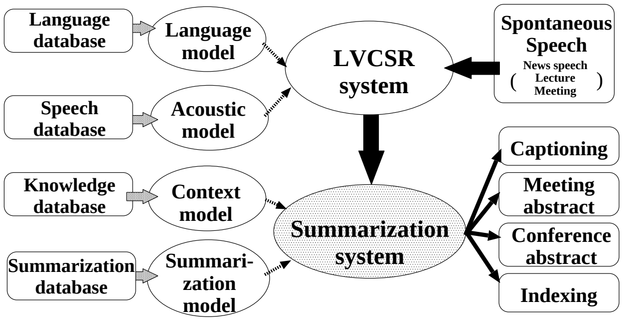


Figure 1: *Automatic speech summarization system.*

Our goal is to build a system that extracts and presents information from spoken utterances according to users' desired amount of information. Figure 1 shows our proposed system. The output of the system can be a summarized sentence for each utterance or summarization of an article consisting of multiple utterances. These outputs can be used for indexing, making closed captions and abstracts, etc. In the closed captioning of broadcast news, the number of words spoken by professional announcers sometimes exceeds the number of words that people can read and understand if all of them are presented on the TV screen. Therefore, reduction of the number of words in speech is indispensable. Meeting/conference summarization should be useful if it can extract relatively important information scattering about in the original speech.

Techniques of automatically summarizing written text have been actively investigated in the field of natural language processing [5]. One of the major techniques for summarizing written text is the process of extracting important sentences. Recently a sentence compression technique using a pair of text and abstract has been proposed [6]. A major difference between text summarization and speech summarization exists in the fact that transcribed speech is sometimes linguistically incorrect due to the spontaneity of speech and recognition errors. A new approach to automatically summarizing speech is needed to cope with such problems.

We have already proposed an automatic speech summarization technique for Japanese speech [7] [8] [9]. Japanese broadcast news speech and lecture speech can be summarized effectively by our proposed method. Since our method is based on a statistical approach, it can be also applied to other languages. In this paper, English broadcast news speech transcribed using a speech recognizer [10] is automatically summarized and its performance is evaluated.

2. SUMMARIZATION OF EACH SENTENCE UTTERANCE

Our method to summarize speech, sentence by sentence, extracts a set of words maximizing a summarization score from an automatically transcribed sentence according to a summarization ratio and concatenates them to build a summary. The summarization ratio is the number of characters/words in the summarized sentence divided by the number of characters/words in the original sentence. The summarization score indicating the appropriateness of a summarized sentence is defined as the sum of a word significance score I, a confidence score C of each word in the original sentence, a linguistic score L of the word string in the summarized sentence [7] [8] and a word concatenation score T [9]. The word concatenation score given by SDCFG indicates a word concatenation probability determined by a dependency structure in the original sentence.

Given a transcription result consisting of N words, $W = w_1, w_2, \ldots, w_N$, the summarization is performed by extracting a set of $M (M < N)$ words, $V = v_1, v_2, \ldots, v_M$, which maximizes the summarization score given by eq.(1).

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

where λ_I, λ_C and λ_T are weighting factors for balancing among L,I, C and T.

This method is effective in reducing the number of words by removing redundant and irrelevant information without losing relatively important information. A set of words maximizing the total score is extracted using a DP technique [7].

2.1 Word significance score

The word significance score I indicates relative significance of each word in a original sentence [7]. The amount of information based on the frequency of each word given by eq. (2) is used as the word significance score for topic words.

$$I(w_i) = f_i \log \frac{F_A}{F_i} \qquad (2)$$

where,

w_i :	a topic word in the transcribed speech
f_i :	number of occurrences of w_i in the transcription
F_i :	number of occurrences of w_i in all the training documents
F_A :	summation of all F_i in all the training documents $(= \sum_i F_i)$

w_i which occurs frequently over all documents is deweighted by our measure given by eq. (2). Our preliminary experiment showed that this measure is more effective than the tf-idf measure in which w_i homogeneously occurring among documents in the collection data is deweighted.

In this study we choose nouns and verbs as topic words for English. A flat score is given to words other than topic words. To reduce the repetition of words in the summarized sentence, a flat score is also given to each reappearing noun and verb.

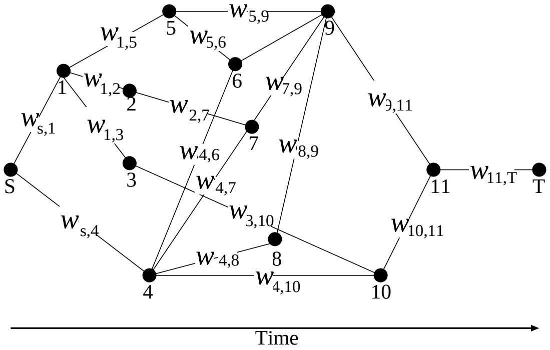


Figure 2: *An example of word graph.*

2.2 Linguistic score

The linguistic score $L(v_m| \ldots v_{m-1})$ indicates the appropriateness of the word strings in a summarized sentence and is measured by n-gram probability $P(v_m| \ldots v_{m-1})$ [7]. In contrast with the word significance score which focuses on topic words, the linguistic score is helpful to extract other words necessary to construct a readable sentence.

2.3 Confidence score

The confidence score $C\ (v_m)$ is incorporated to weight acoustically as well as linguistically reliable hypotheses [8]. Specifically, a posterior probability of each transcribed word, that is the ratio of a word hypothesis probability to that of all other hypotheses, is calculated using a word graph obtained by a decoder and used as a confidence measure [11]. A word graph consisting of nodes and links from a beginning node S to an end node T is shown in Figure 2.

Nodes represent time boundaries between possible word hypotheses and links connecting these nodes represent word hypotheses. Each link is given acoustic log likelihood and linguistic log likelihood of a word hypothesis.

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

$$C(w_{k,l}) = \log \frac{\alpha_k P_{ac}(w_{k,l}) P_{lg}(w_{k,l}) \beta_l}{\mathcal{G}} \quad , \qquad (3)$$

where,

k, l	: node number in a word graph ($k < l$)
$w_{k,l}$	: word hypothesis occurred 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 $w_{k,l}$
$\mathcal{G}$	: forward probability from the beginning node S to the end node $T (= \alpha_T)$

2.4 Word concatenation score

Suppose "the beautiful cherry blossoms in Japan" is summarized as "the beautiful Japan". The latter phrase is grammatically correct but a semantically incorrect summarization. Since the above linguistic score is not powerful enough to alleviate such a problem, a word concatenation score $T(v_{m-1}, v_m)$ is incorporated to give a penalty for a concatenation between words with no dependency in the original sentence. Every language has its own dependency structure and a basic computation of the word concatenation score independent of the type of language is described.

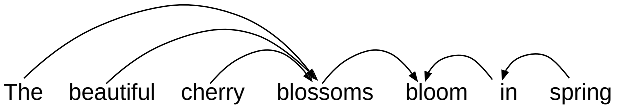


Figure 3: *An example of dependency structure.*

Dependency structure

An example of the dependency structure represented by a dependency grammar is shown as the curved arrows in Figure 3. In a dependency grammar, one word is designated as the head of a sentence, and all other words are either a dependent of that word, or dependent on some other word which connects to the head word through a sequence of dependencies [12]. The word at the beginning of an arrow is named the "modifier" and the word at the end of the arrow is named the "head" respectively. For instance, the dependency grammar of English consists of both "right-headed" dependency indicated by right arrows and "left-headed" dependency indicated by left arrows as shown in Figure 3. These dependencies can be represented by a phrase structure grammar, DCFG (Dependency Context Free Grammar), using the following rewrite rules based on Chomsky normal form.

$$\begin{array}{lll} \alpha & \rightarrow & \beta\alpha \ \ \texttt{(right-headed)} \\ \alpha & \rightarrow & \alpha\beta \ \ \texttt{(left-headed)} \\ \alpha & \rightarrow & w \end{array}$$

where α and β are nonterminal symbols and w is a terminal symbol (word). Figure 4 illustrates an example of a phrase structure tree based on a word-based dependency structure for a sentence which consists of L words, $w_1, \ldots, w_L$. The w_x modifies w_z, when a sentence is derived from the initial symbol S and the following requirements are fulfilled: 1) the rule of $\alpha \rightarrow \beta\alpha$ is applied; 2) $w_i \ldots w_k$ is derived from β; 3) w_x is derived from β; 4) $w_{k+1} \ldots w_j$ is derived from α and 5) w_z is derived from α.

Dependency probability

Since dependencies between words are usually ambiguous, whether dependencies exist or not between words must be estimated by a dependency probability that one word is modified by others. In this study, the dependency probability is calculated as a posterior probability estimated by the Inside-Outside probabilities [13] based on SDCFG. The probability that the w_x and w_z relationship has a "right-headed" dependency structure is calculated as a product of the probabilities of the above-mentioned steps from 1) to 5). On the other hand, the "left-headed" dependency probability is calculated as the product of the probabilities when the rule of $\alpha \rightarrow \alpha\beta$ is applied. Since English has both right and left dependencies, the dependency probability is defined as the sum of the "right-headed" and "left-headed" dependency probabilities. If a language has only "right-headed" dependency, the "right-headed" dependency probability is used for the dependency probability. For simplicity, the dependency probabilities between w_x and w_z is denoted by $d(w_x, w_z, i, k, j)$, where i, k are the indices of the initial and final words derived from β, and j is the index of the final word derived from α. The dependency probability is calculated as follows:

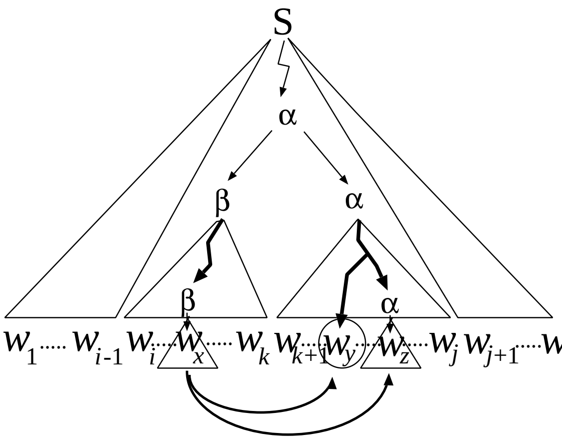

S
α
β
α
β
α
W1 ..... Wi-1 Wi ..... Wx ..... Wk Wk+1 ..... Wy ..... Wz ..... Wj Wj+1 ..... WL

Figure 4: *A phrase structure tree based on a dependency structure.*

$$\begin{aligned} & d(w_m, w_l, i, k, j) \\ & \quad = \Bigg\{ \sum_{\alpha\beta} f(i,j|\alpha) P(\alpha \rightarrow \beta\alpha) h_m(i,k|\beta) h_l(k+1,j|\alpha) \\ & \quad + \sum_{\alpha\beta:\alpha\neq\beta} f(i,j|\alpha) P(\alpha \rightarrow \alpha\beta) h_m(i,k|\alpha) h_l(k+1,j|\beta) \Bigg\} \end{aligned} \tag{4}$$

where P is a rewrite probability and f is outside probabilities given by eq. (A-2) in the Appendix. h is a head-dependent inside probability that w_n is a head of a word string derived from α is defined as follows:

$$\begin{aligned} & h_n(i,j|\alpha) \\ & = \begin{cases} \displaystyle\sum_\beta \Bigg\{ \sum_{k=i}^{n-1} P(\alpha \rightarrow \beta\alpha) e(i,k|\beta) h_n(k+1,j|\alpha) \\ \qquad + \displaystyle\sum_{k=n}^{j-1} P(\alpha \rightarrow \alpha\beta) h_n(i,k|\alpha) e(k+1,j|\beta) \Bigg\} \\ \qquad\qquad\qquad\qquad\qquad\qquad if \ \ i < j \\ b(\alpha \rightarrow w_n) \qquad\qquad\qquad\quad if \ \ i = j = n \\ 0 \qquad\qquad\qquad\qquad\qquad\quad\ otherwise \end{cases} \end{aligned} \tag{5}$$

where e is the inside probability given by eq. (A-1) in the Appendix.

Word concatenation probability

In general, as shown in Figure 4, a modifier derived from β can be directly connected with a head derived from α in a summarized sentence. In addition, the modifier can be also connected with each word which modifies the head. The word concatenation probability between w_x and w_y is defined as a sum of the dependency probabilities between w_x and w_y, and between w_x and each of the $w_{y+1} \ldots w_z$. Using the dependency probabilities $d(w_x, w_y, i, k, j)$, the word concatenation score is calculated as a logarithmic value of the word concatenation probability given by:

$$T(w_x, w_y) = \log \sum_{i=1}^{x} \sum_{k=x}^{y-1} \sum_{j=y}^{L} \sum_{z=y}^{j} d(w_x, w_z, i, k, j). \tag{6}$$

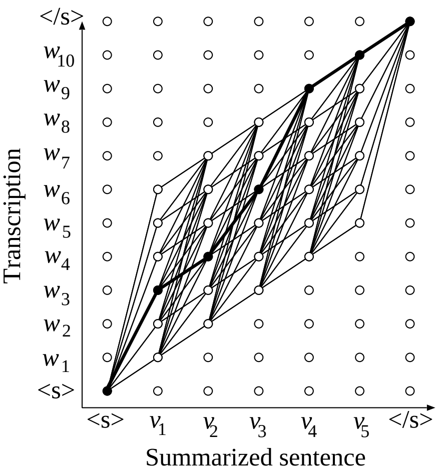


Figure 5: *An example of DP process for speech summarization.*

SDCFG

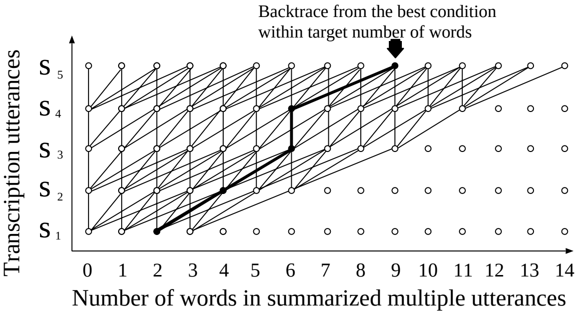
SDCFG is constructed using a manually parsed corpus. Parameters of SDCFG are estimated using the Inside-Outside algorithm. In our SDCFG [14], only the number of non-terminal symbols is determined and all possible phrase trees are considered. The rules consisting of all combinations of non-terminal symbols are applied to each rewriting symbol in a phrase tree. In this method, the non-terminal symbol is not given a specific function such as a noun phrase function, and the function of non-terminal symbols are automatically learned from data. Probabilities for frequently used rules become bigger, and those for rarely used rules become smaller. Since words in the learning data for SDCFG are tagged with POS (part-of-speech), the dependency probability of words excluded in the learning data can be calculated based on their POS. Even if the transcription results obtained by a speech recognizer are ill-formed, the dependency structure can be robustly estimated by the SDCFG.

2.5 Dynamic programming for automatic summarization

Given a transcription result consisting of N words, $W = w_1, w_2, \ldots, w_N$, the summarization is performed by extracting a set of $M (M < N)$ words, $V = v_1, v_2, \ldots, v_M$, which maximizes the summarization score given by eq. (1). The two-dimensional space for performing the dynamic programming process is shown in Figure 5. The vertical axis indicates the transcription result consisting of 10 words ($N = 10$), and the horizontal axis indicates the summarized sentence having 5 words ($M = 5$). All possible sets of 5 words extracted from the 10 words are indicated by the paths from the bottom-left corner to the top-right corner. One of the paths which maximizes the summarization score is selected.

3. SUMMARIZATION OF MULTIPLE UTTERANCES

Our proposed automatic speech summarization technique for each sentence has recently been extended to summarize a set of multiple utterances (sentences) [9]. A set of words maximizing the summarization score is extracted from multiple utterances under some restrictions applied at the sentence boundaries. These restrictions realizes the summarization of multiple utterances by handling them as a single long utterance. This results in preserving more words inside information rich utterances and shortening or even completely deleting less informative ones. This summarization technique can be considered as a combination of the summarization method extracting important sentences investigated in the field of natural language processing and the sentence-by-sentence summarization method. The multiple utterance summarization method should be especially useful for making lecture abstracts, meeting minutes, etc.



Backtrace from the best condition
within target number of words
Transcription utterances
S 5
S 4
S 3
S 2
S 1
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
Number of words in summarized multiple utterances

Figure 6: *An example of DP process for summarization of multiple utterances.*

However, the amount of calculation for selecting the best combination among all possible combinations of words in the multiple utterances increases as the number of words in the original utterances increases. In order to reduce the amount of calculation, we proposes a new method in which each utterance is summarized according to all possible summarization ratio and then the best combination of summarized sentences is determined according to a target compression ratio using a two-level DP technique. Figure 6 illustrates the two-level DP technique for summarizing multiple utterances.

4. EVALUATION

4.1 Word network of manual summarization results for evaluation

In order to automatically evaluate summarized sentences, correctly transcribed speech are manually summarized by human subjects and used as correct targets. The manual summarization results are merged into a word network which approximately expresses all possible correct summarization including subjective variations. A "summarization accuracy" of automatic summarization is calculated using the word network [9]. A word string that is the most similar to the automatic summarization result extracted from the word network is considered as a correct target for the automatic summarization. The accuracy, comparing the summarized sentence with the target word string, is used as a measure of linguistic correctness and maintenance of original meanings of the utterance.

4.2 Evaluation data

English TV broadcast news utterances (CNN news) recorded in 1996 given by NIST as a test set of Topic Detection and Tracking (TDT) were tagged by Brilltagger [16] and used to evaluate our proposed method. Five news articles consisting of 25 utterances in average were transcribed by the JANUS [10] speech recognition system. The multiple utterance summarization was performed for each of the five news articles at 40% and 70% summarization ratio. 50 utterances arbitrarily chosen from the five news articles were used for the sentence by sentence summarization with the summarization ratios of 40% and 70%. Mean word recognition accuracies of the utterances used for the multiple utterance summarization and those for sentence by sentence summarization were 78.4% and 81.4%, respectively. In order to build word networks of manual

Table 1: *Examples of automatic summarization and the corresponding target extracted from a manual summarization word network.* upper: a set of words extracted from the correct summarization network which is the most similar to the automatic summarization, lower: automatic summarization of recognition result.

Recognition result	VICE PRESIDENT AL GORE SAYS THE GOVERNMENT HAS A PLAN TO AVOID THE INEVITABLE PROSPECT OF INCREASED AIRPLANE CRASHES AND FATALITY IS
70% summarization	VICE PRESIDENT AL GORE SAYS THE GOVERNMENT HAS A PLAN TO AVOID THE INCREASED AIRPLANE CRASHES
	VICE PRESIDENT AL GORE SAYS THE GOVERNMENT HAS A PLAN TO AVOID <DEL> INCREASED AIRPLANE CRASHES
40% summarization	<INS> THE GOVERNMENT HAS A PLAN TO AVOID THE INCREASED AIRPLANE CRASHES
	GORE THE GOVERNMENT HAS A PLAN TO AVOID THE INCREASED AIRPLANE CRASHES

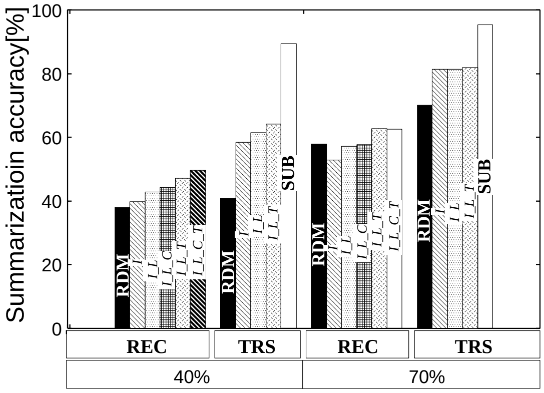


Figure 7: *Each utterance summarization at 40% and 70% summarization ratio.*

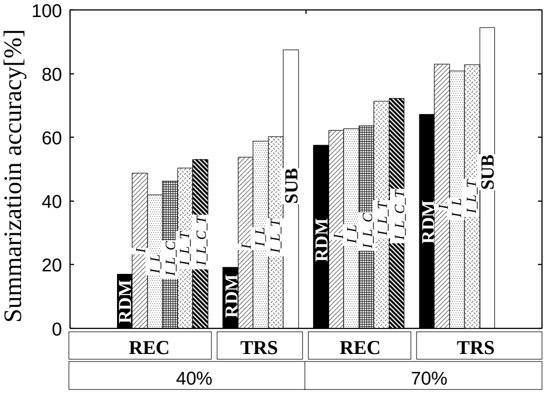


Figure 8: *Article summarization at 30% and 70% summarization ratio.*

summarization results, 17 native English speakers generated manual summarization by removing or extracting words.

4.3 Structure of transcription system

English broadcast news speech was transcribed by the JRTk (Janus Speech Recognition Toolkit) [10] with the following conditions.

Feature extraction

Sounds were digitized with 16kHz sampling and 16bit quantization. Feature vectors had 13 elements consisting of MFCC. Vocal Tract Length Normalization (VTLN) and cluster-based cepstral mean normalization were used to compensate for speaker and channel. Linear Discriminant Analysis (LDA) was applied to reduce feature dimensions in each segment consisting of 7 frames to 42.

Acoustic model

A pentphone model with 6000 distributions sharing 2000 codebooks were used. There were about 105k Gaussians in the system. The training data was comprised of 66 hours of Broadcast News(BN).

Language model

Bigram and trigram were built using BN corpus. Its vocabulary size was 40k.

Decoder

A word-graph-based 3-pass decoder which was composed with JRTk was used for transcription. In the first pass, frame-synchronous beam search was performed using a tree-based lexicon, the above-mentioned HMMs and a bigram model to generate a word graph. In the second pass, frame-synchronous beam search was performed again using a flat lexicon hypothesized in the word graph by the first pass and a trigram model. In the third pass, the word graph was minimized and rescored using the trigram language model.

4.4 Training data for summarization models

A word significance model, a bigram language model and SDCFG were constructed using roughly 35M words (10681 sentences) of the Wall Street Journal corpus and the Brown corpus in Penn Treebank[15].

4.5 Evaluation results

Manual transcription (TRS) and automatic transcription (REC) were both summarized. Table 1 shows examples of automatic summarization and the corresponding target extracted from a manual summarization word network. Figure 7 shows summarization accuracies of utterance summarization at 40% and 70% summarization ratio and Figure 8 shows those of summarizing articles having multiple utterances at 40% and 70% summarization ratio. In these figures, I, L, C and T indicate that the word significance score, the linguistic score, the confidence score and the word concatenation score are used, respectively. In the summarization of REC, conditions with and without the word confidence score, ($I\_L\_C\_T$) and (I_L_T), were compared. In summarization for both TRS and REC, conditions with and without the word concatenation score, (I_L_T, $I\_L\_C\_T$) and (I_L, I_L_C), were compared.

The summarization accuracies for manual summarization (SUB) is considered to be the upper limit of the automatic summarization accuracy. To ensure that our method is sound, we made randomly generated summarized sentences (RDM) according to the summarization ratio and compared them with those obtained by our proposed method.

Table 2: *Number of recognition errors in summarized sentences* **(():** *number of sentences including recognition errors***)**

	Each utterance		Multiple utterances	
REC	180(45)		326(94)	
Summarization ratio	40%	70%	40%	70%
I	42 (27)	111 (40)	99 (56)	199 (71)
I_L	44 (28)	87 (37)	86 (53)	166 (69)
I_L_C	23 (15)	49 (22)	34 (28)	82 (47)
I_L_T	46 (27)	84 (37)	90 (56)	173 (69)
$I\_L\_C\_T$	22 (13)	51 (24)	25 (17)	80 (47)
RDM	82 (30)	87 (21)	89 (45)	169 (65)

These results show that our proposed automatic speech summarization technique is significantly more effective than RDM. By using the word concatenation score (I_L_T, $I\_L\_C\_T$), meaning alteration is reduced compared with the case without using it (I_L, I_L_C). The result obtained when using the word confidence score ($I\_L\_C\_T$) compared with those not using it (I_L_T) shows that the summarization accuracy is improved by the confidence score. Table 2 shows the number of word errors and number of sentences including word errors in the automatic summarization. Recognition errors are effectively reduced by the confidence score.

5. CONCLUSIONS

Each utterance and a whole news article consisting of multiple utterances of English broadcast news speech were summarized by our automatic speech summarization method based on the following scores: word significance score, linguistic likelihood, word confidence measure and word concatenation probability. Experimental results show that our proposed method can effectively extract relatively important information and remove redundant and irrelevant information from English news speech in the same way as for Japanese new speech.

In contrast with the confidence score which has been incorporated into the summarization score to exclude word errors by a recognizer, the linguistic score is effective to reduce out-of-context word extraction both from recognition errors and human disfluencies. In summarizing Japanese news speech, the confidence measure could improve the summarizing performance by excluding in-context word errors. In the English case, the confidence measure can not only exclude word errors but also help extracting clearly pronounced important words. Consequently the use of the confidence measure yields a larger increase in the summarization accuracy for English than Japanese.

6. ACKNOWLEDGMENT

The authors would like to thank Dr. Yoshi Gotoh (Sheffield University) for an arrangement of generating the correct answers for automatic summarization.

7. REFERENCES

[1] T. Imai et al., “Progressive 2-pass Decoder for Real-Time Broadcast News Captioning,” *Proc. ICSLP2000*, vol.I, pp.246–249, Beijing, 2000.

[2] Z. Klaus, “Automatic Generation of Concise Summaries of Spoken Dialogues in Unrestricted Domains,” *Proc. SIGIR2001*, New Orleans, 2001.

[3] S. Furui et al., “Toward the Realization of Spontaneous Speech Recognition -Introduction of a Japanese Priority Program and Preliminary Results-,” *Proc. ICSLP2000*, vol.III, pp.518–521, Beijing, 2000.

[4] R. Valenza et al., “Summarization of Spoken Audio through Information Extraction,” *Proc. ESCA Workshop on Accessing Information in Spoken Audio*, pp.111–116, Cambridge, 1999.

[5] I. Mani et al., ”Advances in Automatic Text Summarization,” *The MIT Press*, 1999.

[6] K. Knight et al., ”Statistics-Based Summarization — Step One: Sentence Compression,” *Proc. National Conference on Artificial Intelligence* (AAAI), 2000.

[7] C. Hori et al., “Automatic Speech Summarization Based on Word Significance and Linguistic Likelihood,” *Proc. ICASSP2000*, vol.III, pp.1579–1582, Istanbul, 2000.

[8] C. Hori et al., “Improvements in Automatic Speech Summarization and Evaluation Methods,” *Proc. ICSLP2000*, vol.IV, pp.326–329, Beijing, 2000.

[9] C. Hori et al., “Advances in Automatic Speech Summarization,” *Proc. EUROSPEECH2001*, vol.III, pp.1771–1774, Aalborg, 2001.

[10] A. Waibel et al., “Advances in Meeting Recognition,” *Proc. HLT2001*, pp.11–13, San Diego, 2001.

[11] T. Kemp et al., “Estimating confidence using word lattices,” *Proc. 5th Eurospeech*, Rhodes, Vol.II, pp.827–830, 1997.

[12] C. Manning et al., “Foundations of Statistical Natural Language Processing,” *The MIT Press*, 2000.

[13] K. Lari et al., “The estimation of stochastic context free grammars using the Inside-Outside algorithm,” *Computer Speech and Language*, 4, pp.35–56, 1990.

[14] A. Ito et al., “Language Modeling by Stochastic Dependency Grammar for Japanese Speech Recognition,” *Proc. ICSLP2000*, Beijing, Vol.I, pp.246–249, 2000.

[15] http://www.cis.upenn.edu/˜treebank/

[16] http://www.cs.jhu.edu/˜brill/

APPENDIX

Inside probability e and outside probablitie f are given by eqs. (A-1) and (A-2), respectively.

Inside Probabilities

$$e(i,j|\alpha) = \begin{cases} \displaystyle\sum_{k=i}^{j-1} \left\{ \sum_{\beta} P(\alpha \to \beta\alpha) e(i,k|\beta) e(k+1,j|\alpha) \right. \\ \qquad \left. + \displaystyle\sum_{\beta:\alpha\neq\beta} P(\alpha \to \alpha\beta) e(i,k|\alpha) e(k+1,j|\beta) \right\} \\ \qquad\qquad\qquad\qquad\qquad if\ \ i<j \\ b(\alpha \to w_i) \qquad\qquad\qquad\quad if\ \ i=j \end{cases} \tag{A-1}$$

Outside Probabilities

$$\begin{aligned} f(i,j|\alpha) \\ = \sum_{k=1}^{i-1} & \left\{ \sum_{\beta} P(\alpha \to \beta\alpha) e(k,i-1|\beta) f(k,j|\alpha) \right. \\ & \left. + \sum_{\beta:\alpha\neq\beta} P(\beta \to \beta\alpha) e(k,i-1|\beta) f(k,j|\alpha) \right\} \\ + \sum_{k=j+1}^{L} & \left\{ \sum_{\beta} P(\beta \to \alpha\beta) e(j+1,k|\beta) f(i,k|\alpha) \right. \\ & \left. + \sum_{\beta:\alpha\neq\beta} P(\alpha \to \alpha\beta) e(j+1,k|\beta) f(i,k|\alpha) \right\} \end{aligned} \tag{A-2}$$