論文 / 著書情報
Article / Book Information

論題(和文)	単語抽出による音声要約文生成法とその評価
Title(English)	
著者(和文)	堀 智織, 古井 貞熙
Authors(English)	SADAOKI FURUI
出典(和文)	電子情報通信学会論文誌, Vol. J85-D-II, No. 2, pp. 200-209
Citation(English)	, Vol. J85-D-II, No. 2, pp. 200-209
発行日 / Pub. date	2002, 2
URL	http://search.ieice.org/
権利情報 / Copyright	本著作物の著作権は電子情報通信学会に帰属します。 Copyright (c) 2002 Institute of Electronics, Information and Communication Engineers.

論　文

単語抽出による音声要約文生成法とその評価

堀　智織[†]　　古井　貞熙[†]

Summarized Speech Sentence Generation Based on Word Extraction and Its Evaluation

Chiori HORI[†] and Sadaoki FURUI[†]

あらまし　本論文では，音声認識結果から発話単位で要約文を生成する音声自動要約手法を提案する．本手法は，原文の文字数を基準とする任意の割合（要約率）を指定して，音声認識結果から単語を抽出し接合することにより要約文を生成する．抽出された部分単語列に対し，要約文としての適正を示す尺度として要約スコアを定義する．この要約スコアを最大とする部分単語列を動的計画法により決定し，自動要約文とする．要約スコアは，要約文中の各単語の単語重要度（重要度スコア），認識時における音響的・言語的信頼度（信頼度スコア），及び要約文内の単語連鎖の言語ゆう度（言語スコア）の累積スコアによって定義される．更に，本論文では提案手法により生成された自動要約文に対し，被験者が単語抽出により作成した正解要約文を基準とする要約文の評価尺度を提案する．すなわち，被験者の作成した正解要約文を単語ネットワークを用いて表現し，ネットワーク上で自動要約文に最も類似している単語列に対し，単語正解精度を要約正解精度として評価する．音声自動要約実験としてNHKのニュース音声を大語彙連続音声認識（LVCSR）システムを用いて音声認識し，20，40，60，70，80%の5段階の要約率で提案手法により自動要約した結果を報告する．更に，その自動要約文を正解要約文単語ネットワークにより評価した結果を示す．実験結果より，自動生成された要約文が，すべての要約率で発話内容を端的に表す重要な情報を保持しつつ，冗長または不要な情報を削減できることを示す．

キーワード　音声要約，要約スコア，動的計画法，正解要約文単語ネットワーク，要約正解精度

1. まえがき

近年の大語彙連続音声認識（LVCSR）技術の進展に伴い，LVCSRシステムを用いて認識された音声を実用的な場面で利用することへの期待が高まっている．現在，LVCSRシステムを用いた放送音声への字幕付与[1]や講演録作成[2]の実用化を目指した検討が行われている．また，音声認識結果から単語抽出により音声データへのインデクシングを行う試み[3]も検討されている．

しかし，LVCSRシステムの認識結果をそのまま字幕や議事録，インデクシング等に用いた場合，冗長な情報が多いという問題がある．そこで，ユーザの使用目的に応じて情報量を制御し，音声の内容を簡潔に表現することが必要である．例えば，放送ニュースの字幕では，ニュースの進行や人間の読解速度に伴う時間的制約により画面表示できる文字数が制限されるため，文字数を圧縮する必要がある．また，会議や講演，講義では，音声全体に渡り散在している重要箇所を，簡潔に抜粋する必要がある．以上の問題は，目的に応じた要約率で音声を要約することにより解決することができる．

これまで自動要約の手法は，テキストを対象として自然言語処理の分野で検討されてきた．テキスト自動要約の手法としては，重要単語や手掛り語に基づき文集合から重要文を抽出し要約文とする手法が主流である[4]．しかし，音声は書き言葉に基づくテキストと異なり，必ずしも文法的に正しいとは限らない．更に，人間の自然発話には，言いよどみや言い直し，「えー」などに代表される間投詞といった冗長な情報が多く含まれている．また，音声認識結果には認識誤りによる不要な情報が含まれる可能性がある．音声認識結果を，字幕，講演録，議事録，インデクシング等に応用する場合，これらの冗長または不要な情報を削除する必要がある．

[†] 東京工業大学大学院情報理工学研究科，東京都
Graduate School of Information Science and Engineering, Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro-ku, Tokyo, 152-8552 Japan

本論文では，単語抽出による音声自動要約手法を提案する．本手法は，音声認識された各発話文からその文字数に対し任意の割合（要約率）で単語を抽出し接合することにより要約文を生成する．ここで，発話文とは，意味のある言葉のまとまりをなす継続的な音声を指すものとし，放送ニュースではアナウンサーの読み上げる各1文を指すものとする．本論文では，この単位を発話単位または単に発話と呼ぶ．

生成された自動要約文は，要約率の範囲で発話文における重要な情報をできる限り保持し，音声認識誤りを含まず，日本語として文意が理解できる文となっていることが求められる．本手法では，これらの条件を考慮して，要約文として抽出される各単語に対し，発話文の情報をどの程度担っているかを表す単語重要度（重要度スコア），音声認識システムによってどの程度確信をもって認識されたかを表す信頼度（信頼度スコア），及び抽出した単語を連接させた際，その単語連鎖がどの程度言語的にもっともらしいかを表す言語ゆう度（言語スコア）を求め，それらを累積したスコアを要約文の適正度を示す尺度とし，要約スコアと定義する．この要約スコアを最大とする要約文を，特定の要約率において抽出し得る複数の要約文候補の中から動的計画法により決定する．

本手法は，要約スコアが最大となる部分単語列を抽出することから，信頼度の低い音声認識誤りを排除しつつ，発話文中の情報の核となる重要単語を重点的に抽出すると同時に，それら重要単語の間を言語的にもっともらしくなるように他の単語で補うことができる．これにより，自動要約文は，単なるキーワード抽出と異なり，重要単語を含んだ「文」として生成される．更に，この音声要約技術は目的に応じて要約率を選択できることから，音声付きの画像への字幕自動付与，議事録や講義・講演などの抄録の自動作成，または音声データや音声データ付き画像データなどのインデクシングに応用できる．また，音声データの情報検索において，重要箇所のみを再生する「ななめ聞き」システムなどにも応用できる．

更に，本論文では，提案手法を適用して生成した自動要約文に対し，被験者が作成した正解要約文を基準とする要約文の定量的評価尺度を提案する．すなわち，正解要約文を単語ネットワークを用いて表現し，ネットワーク上で自動要約文に最も類似している単語列に対する単語正解精度を要約正解精度として評価する．この手法は，被験者間における正解要約文の変動を近似的に網羅することが可能である．

音声自動要約実験としてNHKのニュース音声をLVCSRシステムを用いて音声認識し，20，40，60，70，80%の5段階の要約率で提案手法により自動要約した結果を報告する．更に，その自動要約文を正解要約文単語ネットワークに基づく要約正解精度により評価した結果を示す．実験結果より，自動生成された要約文が，すべての要約率で発話内容を端的に表す重要な情報を保持しつつ，冗長または不要な情報を削減できることを示す．

2. 単語抽出による音声自動要約手法

本論文で提案する音声自動要約手法は，認識された各発話文から発話内容を端的に表す重要な単語を，原文の文字数に対する要約文の文字数の割合（要約率）に関して，ある与えられた要約率で抽出し，それらを接合することにより要約文を生成する．ただし，本論文では形態素を単語と定義する．

本手法は要約文としてのもっともらしさを示す尺度として要約スコアを定義し，この要約スコアを最大とする部分単語列を最適な要約文として動的計画法により決定する．要約スコアは，要約文に抽出された各単語の単語重要度（重要度スコア：I）と要約文内の単語連鎖の言語ゆう度（言語スコア：L），及び認識時における音響的，言語的信頼度（信頼度スコア：C）の累積スコアで定義する．

N 個の単語から成る認識結果 $W = w_1, w_2, \ldots, w_N$ から，要約文として M $(< N)$ 個の単語を抽出し接合した単語列 $V = v_1, v_2, \ldots, v_M$ の要約スコアは，次式によって示される．

$$S(V) = \sum_{m=1}^{M} \{I(v_m) + \lambda_L L(v_m) + \lambda_C C(v_m)\} \tag{1}$$

ただし，λ_L，λ_C は各スコアのバランスをとるための重み係数である．

認識された N 個の単語列 $W = w_1, w_2, \ldots, w_N$ から抽出された部分単語列を $V = v_1, v_2, \ldots, v_M$ $(M < N)$ とするとき，要約処理は式 (1) で表される要約スコアを最大にする $\hat{V}$ を求める問題となり，動的計画法を用いて解くことができる．

2.1 単語重要度スコア

単語重要度スコア $I\,(v_m)$ は，文中における単語の重要度を示すスコアである．本研究では，名詞の単語

重要度スコアとして話題語らしさを示す話題語スコアを適用する．話題語スコアには式 (2) で表される単語の出現頻度に基づく情報量を適用する [5]．

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

w_i：音声認識結果に含まれる名詞
f_i：要約対象である音声中の名詞 w_i の出現頻度
F_i：大規模コーパス中での名詞 w_i の出現頻度
F_A：大規模コーパス中での総名詞数 $(= \sum_i F_i)$

ただし，大規模コーパスは，要約対象と類似したドメインの大量のテキストを集積したコーパスである．本論文では，文献 [5] において被験者の選択した重要単語が主に名詞であったことから，名詞のみを話題語として話題語スコアを定義し，名詞以外の単語には一定値 I_{const} を与えるものとする．ただし，名詞には下記に示すような動作を示す名詞が含まれている．

「検討する」→検討（名詞)+する（動詞)

更に，同一名詞が重複して要約文に出現することを避けるため，2 回目以降の名詞に対しては，名詞以外の単語と同様に一定値 I_{const} を与える．

話題語のみに話題語スコアを用いることにより，話題語以外の単語は言語スコアにより制御され，話題語を含む「文」を構成するために話題語を補う役割を果たす．

2.2 言語スコア

言語スコア $L(v_m)$ は，要約文内の単語連鎖の適正度を示すスコアである．本研究では，統計的言語モデルである単語 trigram を用いる．

$$L(v_m) = \log P(v_m | v_{m-2} v_{m-1}) \qquad (3)$$

2.3 信頼度スコア

信頼度スコア $C(v_m)$ は，認識結果に含まれる認識誤りを要約文に抽出しないよう，音響的，言語的に信頼度の低い単語に対しペナルティを与えるものである．本研究では，デコーダから出力された単語グラフに付与された音響ゆう度及び言語ゆう度に基づく各単語に対する事後確率の対数値を，信頼度スコアとして用いる [6], [7]．図 1 で示すように，単語グラフは文頭ノード S から文末ノード T に至る各ノードとノード間を接続するリンクによって表される．単語間境界を示すノードには時間情報が格納され，単語を示すリンクには各単語の音響ゆう度と言語ゆう度が格納されている．

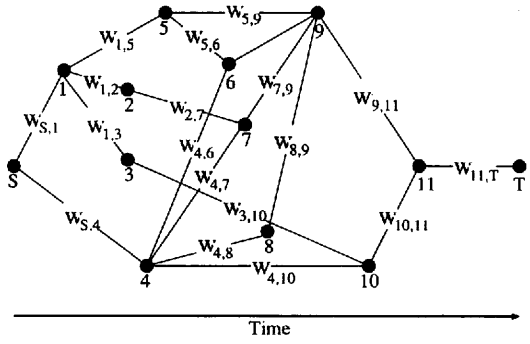

図 1 単語グラフの例
Fig. 1 An example of a word graph.

単語仮説 $\mathrm{w}_{k,l}$ の信頼度スコア $C(\mathrm{w}_{k,l})$ は，次式のように forward 確率と backward 確率を用いた事後確率の対数値として求められる．

$$C(\mathrm{w}_{k,l}) = \log \frac{\alpha_k P_a(\mathrm{w}_{k,l}) P_l(\mathrm{w}_{k,l}) \beta_l}{\mathcal{G}} \qquad (4)$$

k, l：単語グラフにおけるノード番号 $(k < l)$
$\mathrm{w}_{k,l}$：ノード k, l 間のリンクに対応する単語
α_k：始端 S からノード k までの forward 確率
β_l：ノード l から終端 T までの backward 確率
$P_a(\mathrm{w}_{k,l})$：単語 $\mathrm{w}_{k,l}$ の音響ゆう度
$P_l(\mathrm{w}_{k,l})$：単語 $\mathrm{w}_{k,l}$ の言語ゆう度
$\mathcal{G}$：始端 S から終端 T までの forward 確率

この信頼度スコアは，認識された各単語と単語グラフにおける対立候補のゆう度比を示す値であり，値が大きいほど高い信頼度で認識されたとみなすことができる．

2.4 動的計画法に基づく音声要約

N 単語からなる認識結果の単語列 $W = w_1, w_2, \ldots, w_N$ から，$M(< N)$ 単語からなる単語列 $V = v_1, v_2, \ldots, v_M$ を抽出し，要約文候補の中から式 (1) で与えられる要約スコアを最大化する要約文を決定するアルゴリズムを以下に示す．

（1） 記号と変数の定義

<s>：文頭記号
</s>：文末記号
$L(w_n)$：言語スコア
$I(w_n)$：重要度スコア
$C(w_n)$：信頼度スコア
$s(n)$：各単語の要約スコア

$$s(n) = I(w_n) + \lambda_L L(w_n) + \lambda_C C(w_n)$$

ただし，本研究では言語スコアとして trigram を用いていることから，単語 w_n の要約スコアは先行2単語 $w_k w_l$ を考慮して，次式のように定義する．

$$s(k,l,n) = I(w_n) + \lambda_L \log P(w_n|w_k w_l) + \lambda_C C(w_n)$$

$g(m,l,n)$：局所最適スコア

m 単語で構成され，<s> で始まり w_l, w_n で終わる部分単語列 <s>, ..., w_l, w_n の要約スコア，ただし，$(0 \leqq l < n \leqq N)$

$B(m,l,n)$：バックポインタ

（2） 初期設定

$$g(1,0,n) = \begin{cases} I(w_n) + \lambda_L \log P(w_n|\text{<s>}) + \lambda_C C(w_n) \\ \qquad \text{if } 1 \leqq n \leqq (N-M+1) \\ -\infty \quad \text{otherwise} \end{cases}$$

（3） 漸化式計算

for $m = 2$ to M

for $n = m$ to $N - m + 1$

for $l = m - 1$ to $n - 1$

$$g(m,l,n) = \max_{k<l}\{g(m-1,k,l) + s(k,l,n)\}$$

$$B(m,l,n) = \operatorname*{argmax}_{k<l}\{g(m-1,k,l) + s(k,l,n)\}$$

（4） 最適パスの選択

$$S(\hat{V}) = \max_{\substack{N-M<n\leqq N \\ N-M-1<l\leqq N-1}} g(M,l,n) + \log P(\text{</s>}|w_l w_n)$$

$$(\hat{n},\hat{l}) = \operatorname*{argmax}_{\substack{N-M<n\leqq N \\ N-M-1<l\leqq N-1}} g(M,l,n) + \log P(\text{</s>}|w_l w_n)$$

（5） トレースバック

for $m = M$ to 1

$v_m = w_{\hat{n}}$

$l' = B(m,\hat{l},\hat{n})$

$\hat{n} = \hat{l}$

$\hat{l} = l'$

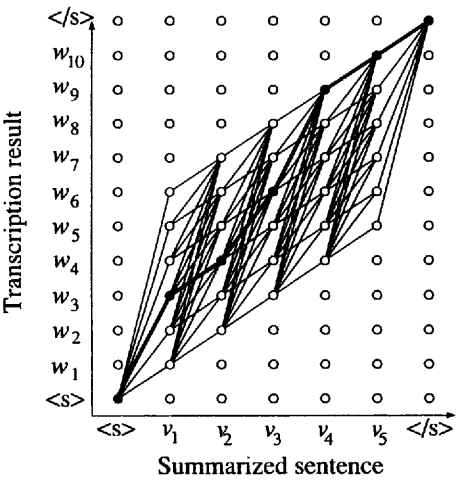

図2 音声要約のための動的計画法の計算領域 ($N = 10, M = 5$)

Fig. 2 An example of DP alignment for speech summarization. ($N = 10, M = 5$)

動的計画法の処理過程の例を図2に示す．縦軸は，音声認識結果により得られた単語列を示し，横軸は要約文として抽出された部分単語列を示す．文頭から文末に向い，音声認識結果の単語列から抽出し得る部分単語列のすべての組合せを，2次元空間に示す．図2の例では，要約スコアが最大となる部分単語列 $w_3, w_4, w_6, w_9, w_{10}$ が，要約文 $v_1, \ldots, v_5$ として抽出される．

ただし，この手法は単語数を基準に要約文を生成しているが，与えられる要約率は文字数が基準である．しかし，要約文の文字数は抽出された単語によって変動するため，目標の要約率に対応した単語数 M を特定することはできない．そこで，単語数 M を変化させた複数の要約文を生成し，与えられた要約率以下で文字数最大となる要約文を選択する．

3. 自動生成要約文に対する評価尺度

3.1 正解要約文ネットワークに基づく要約正解精度

被験者の作成した正解要約文は，被験者により単語の組合せが異なる．そこで，正解要約文の単語間の連鎖をネットワークにまとめることにより，すべての可能性のある正解要約文の単語連鎖を近似的に網羅する．本研究では，正解要約文単語ネットワーク上の単語列を正解として評価する要約正解精度を提案する．5人の被験者による正解要約文の例を表1に示す．更に，その正解要約文に基づく正解要約文単語ネットワーク

表1 被験者による正解要約文の例
Table 1 An exmaple of manual summarization results by human subjects.

要約対象	<s> 日本 で 開かれている 地球 温暖化 国際 会議 </s>
被験者 A	<s> 日本 で 開かれている ___ ___ ___ 会議 </s>
被験者 B	<s> 日本 で 開かれている ___ ___ 国際 会議 </s>
被験者 C	<s> 日本 で 開かれている ___ 温暖化 国際 会議 </s>
被験者 D	<s> 日本 で 開かれている ___ 温暖化 国際 会議 </s>
被験者 E	<s> 日本 で ___ ___ 温暖化 国際 会議 </s>
被験者 F	<s> ___ ___ ___ 地球 温暖化 ___ 会議 </s>

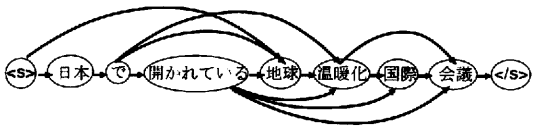

図3 正解要約文単語ネットワークの例
Fig. 3 An example of a word network expressing manual summarizaion results.

を図3に示す．ただし，単語ネットワーク上の右矢印は単語間の連接を示している．このネットワーク上で，文頭記号<s>から文末記号</s>まで右矢印で結ばれた単語列は，要約文として重要な情報を保持し，言語的に正しく，原文の文意を保持した単語連鎖であることから，正解要約文と仮定できる．

そこで，正解要約文ネットワーク上で自動要約文自身に最も近い単語列を正解として，式(5)で表される要約正解精度により自動要約文を評価する．

$$Sum\_acc = \frac{Len - Sub - Ins - Del}{Len} \times 100[\%] \tag{5}$$

Sum_acc: 要約正解精度
Sub: 置換誤り
Ins: 挿入誤り
Del: 削除誤り
Len: ネットワーク上で最も類似している単語列における単語数

図3で示した例では，「<s> 日本 で 会議 </s>」という自動要約文は，ネットワーク上の「<s> 日本 で 温暖化 会議 </s>」という単語列に最も類似している．この両者を比較すると，削除誤りが1であることから，この自動要約文の要約正解精度は75%となる．

要約正解精度は，正解要約文単語ネットワークを用いることにより，重要な情報の抽出，文意の保持，要約文としての適正について，まとめて一元的に評価できる特徴がある．ただし，この尺度は，自動要約文が人間の作成し得る要約文にパターンとしてどの程度類似しているかを定量的に示すものであり，人間が要約文の良さを感覚的に評価した結果と完全に一致するものではない．しかしながら，人間は単語列から文の内容を理解することを考慮すると，強い相関があるものと考えられる．また，人間による感覚的な評価は，被験者の心的状態に左右されることから，適切な評価結果を得るには多数の被験者に様々な条件下で生成された大量の要約文の評価を依頼しなければならない．それに比べると本評価法は簡易で定量的な評価法として有効と考えられる．

4. 評価実験

4.1 実験条件

NHKの女性アナウンサーによるニュース音声（419発話）を音声認識し，そのうち，単語正解精度が90%以上の50発話を要約対象とした．ニュース音声は，事前に人手により発話単位に切り出して用いた．また，この50発話には，言い直しはなく，言いよどみ1個と間投詞13個が含まれている．この発話ごとの音声認識結果に対して，提案する音声自動要約手法を用い，20，40，60，70，80%の5段階の要約率で要約文を生成した．

信頼度スコア（C），単語重要度スコア（I），言語スコア（L），を単独に用いた場合，二つのスコアを組み合わせた場合（I_C, L_C, I_L），すべてのスコアを用いた場合（I_L_C）の全7種類の自動要約文を生成した．各自動要約文は，25人の被験者によって作成された正解要約文に基づき，要約正解精度により評価した．ただし，言語スコア及び信頼度スコアの重み係数 λ_L，λ_C と単語重要度スコアにおける一定値 I_{const} は，様々に変化させ求められた最適値を用いた．更に，音声認識の精度が100%であった場合の音声自動要約の性能を調べるため，人手により音声を書き起こした正解書き起こし文について自動要約を行った．また，正解書き起こし文に対する各被験者の要約文については，その他の24人の被験者の要約文を正解として評価し（SUB），自動要約文の目標値とした．提案手法の有効性を示すため，無作為に単語抽出した要約文（RDM）を生成し，比較対象とした．

4.2 音声認識システムの構成

[特徴抽出]

音声データを16 kHz，16 bitでディジタル化し，フレーム長25 ms，フレーム周期10 msで Δ 対数パワー

と12次元のメルケプストラム及びΔメルケプストラム（計25次元）を抽出する．更に発話ごとにケプストラム平均正規化を行う．

[音響モデル]

8混合ガウス分布，1012状態，不特定話者音素文脈依存HMMのIPAの女性モデル[8]を用い，評価話者の985発話（約2時間）を用いて最ゆう推定による話者適応を行った．

[言語モデル]

単語bigram，trigramを用いる．放送ニュース原稿テキスト5年分（1992年7月から1996年5月）の約50万文を，形態素解析システムJUMAN[9]を用いて形態素に分解し，「単語＋読み＋品詞」を一つの形態素としてbigramとtrigramを学習した．

[デコーダ]

単語グラフを中間表現とする2パスデコーダ[10]を用いる．第1パスではHMMとbigramを用いてフレーム同期のビームサーチを行い，単語グラフを生成する．このとき，単語間の音素文脈依存も考慮する．第2パスでは，単語グラフとtrigramを用いてリスコアを行い，1ベストを認識結果として自動要約処理部へ受け渡す．

4.3 要約処理部の構成

[単語重要度スコア]

音声認識用言語モデルを学習した放送ニュース原稿テキスト5年分（1992年7月から1996年5月）の約50万文を用い，全文における各単語の出現頻度に基づき重要度スコアを求める．

[要約用言語モデル]

要約用言語モデルは要約文における単語連鎖をモデル化したものであるが，言語モデルを学習できる要約文の大規模なコーパスは存在していない．そこで，要約対象であるニュース音声と話題が重なっており，要約文に求められる簡潔な表現（少ない修飾語句，助詞の省略，体言止め等）が多く含まれている新聞記事テキストを要約用言語モデルの学習に用いた．具体的には，3年分の毎日新聞（1996～1998）のテキストを用い，音声認識に用いた言語モデルと同一の2万語彙からなる単語trigramを学習した．

以下に示すように，言語モデルの作成に用いたニュース原稿と新聞記事の1文当りの平均形態素数を比べると，明らかに新聞記事の方が少なく，新聞記事が単文で構成される端的な表現を数多く含むことを裏づけている．

ニュース原稿：44形態素/文
新聞記事　　：17形態素/文

[信頼度スコア]

音声認識システムの第1パスから得られた単語グラフ上で，1ベストの認識結果の各単語について音響ゆう度と言語ゆう度に基づき事後確率の対数値を求める．

5. 実験結果

実験結果の例を表2に示す．実験結果より，要約率に従い重要な情報が保持され，日本語としてもっともらしい要約文が作成されていることがわかる．更に，LVCSRシステムによる音声認識結果を対象とした自動要約文の例では，音声認識誤りが要約文に含まれることが回避できている．要約対象である50発話の音声認識結果には，音声認識誤りのうち「等」が「など」と仮名に置き換わる等の文意に影響のない誤りを除き，69単語の誤りが含まれている．これらの認識誤りが要約文に抽出された場合，発話意図とは異なる要約文が生成されてしまう．提案手法による自動要約文に抽出された誤認識単語の数と，その誤認識単語が含まれる自動要約文の数を表3に示す．表3より，提案手法が誤認識された単語が含まれていることによる誤要約を削減できていることがわかる．

6. 要約正解精度による評価結果と考察

正解要約文単語ネットワーク上で最も類似している正解単語列と自動要約文を比較した例を表4に示す．ただし，各要約率の上段が正解要約文ネットワーク上で自動要約結果に最も類似していた単語列，下段は自動要約結果を示す．

更に，図4から図8に，要約正解精度による評価結果を示す．すべての自動要約条件において，ランダムに単語を抽出した場合と比較して，要約正解精度が高くなることが示された．音声認識を対象とした自動要約文（REC）では，各スコアを単独で適用した場合，いずれの要約率においても言語スコア（L）による要約精度の改善が最も大きいことがわかる．また，すべての自動要約条件を相互に比較した場合，要約率40%以下では，単語重要度スコアと言語スコアの組合せ（I_L）が，要約率60%以上ではすべてのスコアの組合せ（I_L_C）による要約正解精度が最も高いことが示されている．音声認識結果の単語正解精度が90%以上であったことから，信頼度スコアの効果は認識誤り

表2 書き起こしと音声認識結果に対する自動要約結果
Table 2 Summarization results for manual and automatic transcription.

書き起こし	ジュネーブで開かれている地球温暖化対策の国際会議で日本政府は西暦二千年以降先進各国が GDP 国内総生産に応じた二酸化炭素の排出削減に努めるという新たな国際目標を日本としては今回初めて提案することを決めました
要約率 80%	ジュネーブで開かれている地球温暖化対策の国際会議で日本政府は＿＿＿＿先進各国が＿＿＿ 国内総生産に＿＿二酸化炭素の排出削減＿＿＿＿＿＿＿＿目標を日本としては今回初めて提案することを決めました
要約率 70%	ジュネーブで開かれている地球温暖化対策の国際会議で日本政府は＿＿＿＿先進各国が＿＿＿ ＿＿＿＿二酸化炭素の排出削減＿＿＿＿＿＿＿＿目標を日本として＿＿＿＿＿提案することを決めました
要約率 60%	ジュネーブで開かれている地球温暖化対策の国際会議で日本政府は＿＿＿＿＿＿＿＿ ＿＿＿＿＿二酸化炭素の排出削減＿＿＿＿＿＿＿＿目標を日本として＿＿＿＿＿提案することを決めました
要約率 40%	＿＿＿＿＿＿＿＿地球温暖化対策＿＿＿会議で日本政府は＿＿＿＿＿＿＿＿ ＿＿＿＿＿二酸化炭素の排出削減＿＿＿＿＿＿＿＿目標を＿＿＿＿＿＿＿＿提案することを決め＿＿＿
要約率 20%	＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿ ＿＿＿＿＿二酸化炭素の排出削減＿＿＿＿＿＿＿＿目標を＿＿＿＿＿＿＿＿提案すること＿＿＿＿＿
音声認識結果	<年>で開かれている<月いう>温暖化対策の国際会議で日本政府は西暦二千年以降先進各国が GDP 国内総生産に応じた二酸化炭素の排出削減に努めるという新たな国際目標を日本としては今回初めて提案することを決めました
要約率 80%	＿＿＿＿＿＿＿＿＿＿温暖化対策の国際会議で日本政府は＿＿＿＿＿先進各国が＿＿＿＿＿ ＿＿＿＿＿二酸化炭素の排出削減に努めるという新たな国際目標を日本としては今回初めて提案することを決めました
要約率 70%	＿＿＿＿＿＿＿＿＿＿温暖化対策の国際会議で日本政府は＿＿＿＿＿＿先進各国が＿＿＿＿＿ ＿＿＿＿＿二酸化炭素の排出削減に努めるという＿＿＿＿＿＿日本としては今回初めて提案することを決めました
要約率 60%	＿＿＿＿＿＿＿＿＿＿温暖化対策の国際会議で日本政府は＿＿＿＿＿先進各国が＿＿＿＿＿ ＿＿＿＿＿二酸化炭素の排出削減に努めるという＿＿＿＿目標＿日本として＿＿＿＿＿提案することを決めました
要約率 40%	＿＿＿＿＿＿＿＿＿＿温暖化対策の国際会議で日本＿＿＿＿＿＿＿＿＿＿＿＿＿＿ ＿＿＿＿＿二酸化炭素の排出削減＿＿＿＿＿＿＿＿目標を＿＿＿＿＿＿＿＿提案することを決めました
要約率 20%	＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿ ＿＿＿＿＿二酸化炭素の排出削減＿＿＿＿＿＿＿＿目標＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

＿は削除された領域，<>は認識誤りを表す

表3 正解要約文に含まれる認識誤りと認識誤りが含まれる要約文の数
Table 3 Number of word errors and summarized sentences including word errors.

	RDM	自動要約文
要約前	69 単語 (50 文)	
80%	36 単語 (17 文)	12 単語 (8 文)
70%	31 単語 (16 文)	5 単語 (5 文)
60%	25 単語 (15 文)	3 単語 (3 文)
40%	18 単語 (13 文)	2 単語 (2 文)
20%	8 単語 (7 文)	3 単語 (3 文)

() 内は認識誤りが含まれる要約文の数を表す．

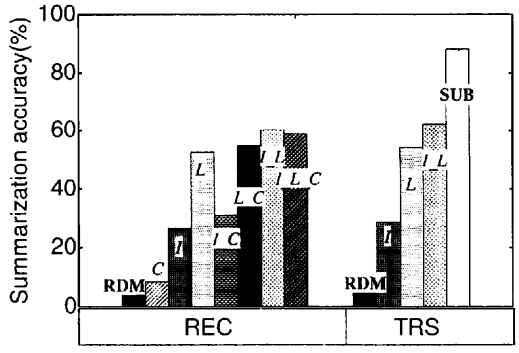

図4 要約率 20%時における要約正解精度．REC：音声認識の自動要約文，TRS：正解書き起こし文の自動要約文，RDM：無作為単語抽出，C：信頼度スコア，I：単語重要度スコア，L：言語スコア，I_C,L_C,I_L：上記二つのスコアの組合せ，I_L_C：上記全スコアの組合せ，SUB：被験者平均

Fig. 4 Summarization results at 20% summarization ratio. REC: summarization of recognition results, TRS: summarization of manual transcription, RDM: random word selection, C: confidence score, I: significance score, L: linguistic score, I_C,L_C,I_L: combination of 2 scores, I_L_C: combination of all scores, SUB: subjective summarization.

が要約文に抽出される可能性の高い高要約率の場合に限られている．

一方，認識誤りを含まない正解書き起こし文に対する自動要約文（TRS）でも，各スコアを単独で適用した場合，いずれの要約率においても言語スコア（L）による要約精度の改善が大きく，単語重要度スコアと言語スコアの組合せ（I_L）の要約精度が最大である．しかし，被験者の作成した正解要約文の要約正解精度には至っていない．

音声認識を対象とした自動要約文（REC），及び正解書き起こし文に対する自動要約文（TRS）のどちらにおいても，単語重要度スコア（I）よりも言語スコア（L）による改善が大きい．被験者の作成した正解要約文では，話題語を中心として，文意を保持する形でそれ以外の単語が抽出されることにより文が構成されている．単語重要度スコア I の重みが大きい場合，

表4　正解要約文単語ネットワークによる要約文評価の例
Table 4　An example of evaluation results based on a manual summarization word network.

音声認識結果	＜年＞で開かれている＜月いう＞温暖化対策の国際会議で日本政府は西暦二千年以降先進各国が GDP 国内総生産に応じた二酸化炭素の排出削減に努めるという新たな国際目標を日本としては今回初めて提案することを決めました
要約率 80%	地球 温暖 化 対策 の 国際 会議 で 日本 政府 は GDP に 応じた 二酸化炭素 の 排出 削減 に 努める と いう 新たな 国際 目標 を 日本 として は 今回 初めて 提案 する こと を 決め ました
	DEL 温暖 化 対策 の 国際 会議 で 日本 政府 は <先進> <各国> <が> 二酸化炭素 の 排出 削減 に 努める と いう 新たな 国際 目標 を 日本 として は 今回 初めて 提案 する こと を 決め ました
要約率 70%	温暖 化 対策 の 国際 会議 で 日本 政府 は GDP に 応じた 二酸化炭素 の 排出 削減 に 努める と いう 目標 を 日本 として は 今回 初めて 提案 する こと を 決め ました
	温暖 化 対策 の 国際 会議 で 日本 政府 は <先進> <各国> <が> 二酸化炭素 の 排出 削減 に 努める と いう DEL DEL 日本 として は 今回 初めて 提案 する こと を 決め ました
要約率 60%	温暖 化 対策 の 国際 会議 で 日本 政府 は 先進 各国 が二酸化炭素 の 排出 削減 に 努める と いう 目標 INS INS INS 提案 する こと を 決め ました
	温暖 化 対策 の 国際 会議 で 日本 政府 は 先進 各国 が二酸化炭素 の 排出 削減 に 努める と いう 目標 日本 と して 提案 する こと を 決め ました
要約率 40%	温暖 化 対策 の 国際 会議 で 日本 政府 二酸化炭素 の 排出 削減 INS を 提案 する こと を 決め ました
	温暖 化 対策 の 国際 会議 で 日本 DEL 二酸化炭素 の 排出 削減 目標 を 提案 する こと を 決め ました
要約率 20%	二酸化炭素 の 排出 削減 目標 提案
	二酸化炭素 の 排出 削減 目標 <DEL>

上段：正解要約文，下段：自動要約文，<>は置換，INS は挿入，DEL は脱落を表す

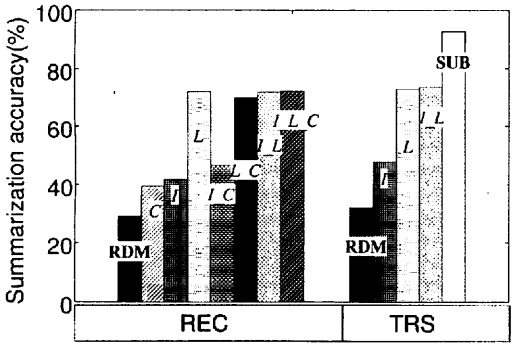

図5　要約率 40%時における要約正解精度．C：信頼度スコア，I：単語重要度スコア，L：言語スコア，I_C,L_C,I_L：上記二つのスコアの組合せ，I_L_C：上記全スコアの組合せ

Fig. 5　Summarization results at 40% summarization ratio. C: confidence score, I: significance score, L: linguistic score. I_C,L_C,I_L: combination of 2 scores, I_L_C: combination of all scores.

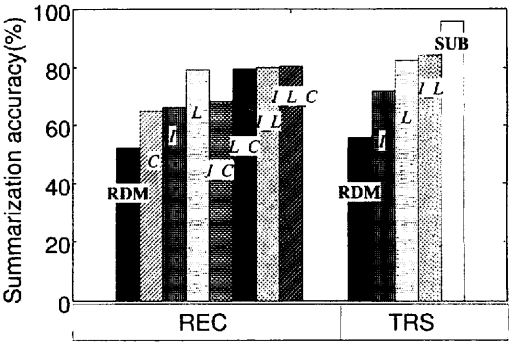

図6　要約率 60%時における要約正解精度．C：信頼度スコア，I：単語重要度スコア，L：言語スコア，I_C,L_C,I_L：上記二つのスコアの組合せ，I_L_C：上記全スコアの組合せ

Fig. 6　Summarization results at 60% summarization ratio. C: confidence score, I: significance score, L: linguistic score, I_C,L_C,I_L: combination of 2 scores, I_L_C: combination of all scores.

話題語のみが抽出されることにより，文としての形態が整わず，要約精度が低下してしまう．一方，言語スコア L の重みが大きい場合，単語重要度のみが考慮された場合に比べ，文として形態が整った自動要約文が生成される．その結果，単語重要度スコア（I）を単独に用いた場合に比べ，言語スコア（L）を単独に用いた自動要約文の要約正解精度がより高くなっているものと考えられる．更に，単語重要度スコアと言語スコアを組み合わせた場合（I_L），正解要約文に近い形で，話題語を中心とした文を構成することができるため，最も高い要約正解精度が得られたものと考えられる．

また，単語抽出に近い低要約率の自動要約文で，自動要約文の性能が劣化している．主な原因は，要約率の低下に伴い，発話内容を端的に表す単語をより的確に選択する必要があるが，単語重要度スコアが不完全なため，単語の重要度の順位が必ずしも正しくないことによるものと考えられる．これは，単語重要度スコ

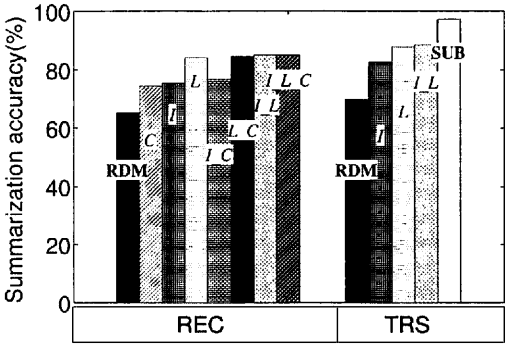

図 7 要約率 70%時における要約正解精度．C：信頼度スコア，I：単語重要度スコア，L：言語スコア，I_C,L_C,I_L：上記二つのスコアの組合せ，I_L_C：上記全スコアの組合せ

Fig. 7 Summarization results at 70% summarization ratio. C: confidence score, I: significance score, L: linguistic score, I_C,L_C,I_L: combination of 2 scores, I_L_C: combination of all scores.

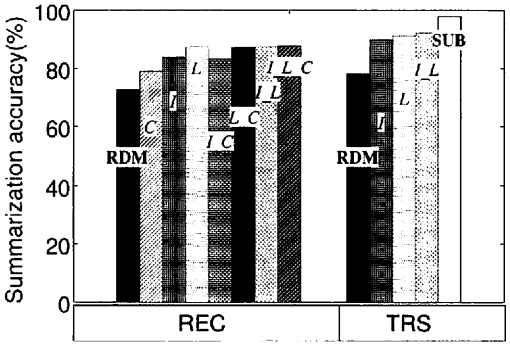

図 8 要約率 80%時における要約正解精度．C：信頼度スコア，I：単語重要度スコア，L：言語スコア，I_C,L_C,I_L：上記二つのスコアの組合せ，I_L_C：上記全スコアの組合せ

Fig. 8 Summarization results at 80% summarization ratio. C: confidence score, I: significance score, L: linguistic score, I_C,L_C,I_L: combination of 2 scores, I_L_C: combination of all scores.

アの評価セットに対する適応等により改善できるものと考えられる．

7. むすび

本論文では，音声自動要約手法として単語重要度スコア（単語重要度），言語スコア（言語ゆう度），及び信頼度スコア（信頼尺度）に基づき，任意の要約率で音声認識結果から動的計画法により単語抽出し，要約文を生成する手法を提案した．更に，自動要約文の評価方法として，正解要約文単語ネットワークに基づく要約正解精度による評価法を提案した．実験結果より，音声認識結果に含まれる認識誤りによる誤要約を抑制しつつ，話題語を中心として日本語としてもっともらしい要約文を生成できることを示した．

今後は，本手法を，より長い発話単位（段落単位または講演等）を対象とした抄録作成，及び，精度の低い認識結果からの重要かつ信頼性の高い部分の抽出といったより実用的な問題に対して適用していく予定である．

また，本研究で用いた被験者の作成した正解要約文との比較による評価では，すべての正解の可能性を網羅することが必ずしもできないという問題がある．更に，日本語として文法は完全には満たされていないが，文意は理解できる要約文については考慮されていない．今後は，原文の文意の保持という点からタスクに応じた評価を組み合わせる必要がある．

謝辞 放送ニュースのデータベースを提供してくださったNHK放送技術研究所に感謝致します．京大コーパスを提供してくださった京大言語メディア研究室に感謝致します．

文　献

[1] 今井　亨，小林彰夫，佐藤庄衛，安藤彰男，“逐次2パスデコーダを用いたニュース音声認識システム，” 信学技報，SP99-129，1999.

[2] 篠崎隆宏，斎藤洋平，堀　智織，古井貞熙，“話し言葉音声の認識を目指して，” 信学技報，SP2000-96，2000.

[3] R. Valenza, T. Robinson, M. Hickey, and R. Tucker, “Summarization of spoken audio through information extraction,” Proc. ESCA Workshop on Accessing Information in Spoken Audio, pp.111-116, 2000.

[4] I. Manu and M. Maubury, Advances in Automatic Text Summarization, The MIT Press, 1999.

[5] 岩崎　淳，古井貞熙，“ニュース音声からの話題抽出法の検討，” 音響学秋季講論，1-1-14，1998.

[6] T. Kemp and T. Schaaf, “Estimating confidence using word lattices,” Proc. 5th Eurospeech, Rhodes, vol.2, pp.827-830, 1997.

[7] V. Valtchev, J.J. Odel, P.C. Woodland, and S.J. Young, “MMIE training of large vocabulary recognition systems,” Speech Communication, vol.22, pp.303-314, 1997.

[8] “日本語ディクテーション基本ソフトウェア，” IPA（情報処理振興事業協会），http://www.lang.astem.or.jp/dictation-tk/

[9] S. Kurohashi, T. Nakamura, Y. Matsumoto, and M. Nagao, “Improvements of Japanese morphological analyzer JUMAN,” Proc. Int. Workshop on Sharable Natural Language Resources, Nara, Aug. 1994.

[10] 堀　貴明，岡　直生，加藤正治，伊藤彰則，好田正紀，“大語彙連続音声認識のための音素グラフに基づく仮説制限法の検討，” 情処学論，vol.40，no.4，pp.1365–1373，1999.

（平成 13 年 4 月 26 日受付，8 月 21 日再受付）

堀　智織

平 6 山形大・工・電子情報卒．平 9 同大大学院博士前期課程了．同年山形大・人文・助手．平 11 同大退官．現在，東工大大学院博士後期課程在学中．音声認識の研究に従事．日本音響学会会員．

古井　貞熙　（正員）

昭 43 東大・工・計数卒．昭 45 同大大学院修士課程了．同年 NTT 電気通信研究所入社．以後，同研究所において，音声認識，話者認識，音声知覚などの研究に従事．昭 53～54 ベル研究所客員研究員．昭 61 NTT 基礎研究所第四研究室長．平 1 NTT ヒューマンインタフェース研究所音声情報研究部長．平 3 同研究所古井特別研究室長．平 6 東京工業大学客員教授．平 9 東京工業大学大学院情報理工学研究科計算工学専攻教授．工博．平 1 科学技術庁長官賞，IEEE ASSP Society Senior Award 受賞．本会より，昭 50 米沢賞，昭 63，平 5 論文賞，平 2 著述賞受賞．日本音響学会より，昭 60，62 佐藤論文賞など受賞．平 8 IEEE Signal Processing Society Distinguished Lecturer．著書「ディジタル音声処理」，「Digital Speech Processing, Synthesis, and Recognition」，「音響・音声工学」，「音声情報処理」，編著「Advances in Speech Signal Processing」など．IEEE 及び米国音響学会（ASA）Fellow．Journal of Speech Communication の Chief Editor．日本音響学会会長．