音声言語情報処理 44－38
(2002. 12. 20)

重要文抽出と文圧縮による音声自動要約

菊池 智紀† 古井 貞熙† 堀 智織††

† 東京工業大学大学院 情報理工学研究科 計算工学専攻
〒 152-8552 東京都目黒区大岡山 2-12-1
†† NTT コミュニケーション基礎科学研究所 知能情報部
〒 619-0237 京都府相楽群精華町光台 2-4
E-mail: †{kikuchi,furui}@furui.cs.titech.ac.jp, ††chiori@cslab.kecl.ntt.co.jp

あらまし 本稿では，これまで我々が提案してきた単語抽出による要約手法の前処理として，重要文抽出を組み合わせた 2 段階の音声自動要約手法を提案する．本手法では音声認識の結果から，各文の構成単語の重要度，信頼度，言語的自然さの評価値から重要文抽出の要約スコアを求め，それをもとに認識率の低い文，理解困難な文をあらかじめ除いておく．次に，残された文に対して，同様の評価値に単語間遷移スコアを加えた要約スコアを最大にするような，部分単語列を抽出するという手法により要約文を作成し，高精度化をはかる．この手法を用いて講演音声を自動要約し，複数の被験者により作成された正解要約文単語ネットワークに基づく評価を行う．重要文抽出法を用いない従来までの要約手法との要約精度の比較を行った結果，提案手法の有効性が確認された．
キーワード 話し言葉，講演音声，音声自動要約，重要文抽出，単語抽出

Automatic speech summarization based on sentence extraction and compaction

Tomonori KIKUCHI†, Sadaoki FURUI†, and Chiori HORI††

† Department of Computer Science, Tokyo Institute of Technology
2-12-1 Ookayama, Meguro-ku, Tokyo, 152-8552 Japan
†† NTT Communication Science Laboratories
2-4, Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0237 Japan
E-mail: †{kikuchi,furui}@furui.cs.titech.ac.jp, ††chiori@cslab.kecl.ntt.co.jp

Abstract This paper proposes a new automatic speech summarization method having two stages: important sentence extraction and sentence compaction. Relatively important sentences are extracted based on the amount of information and the confidence measures of constituent words, and the set of extracted sentences is compressed by our sentence compaction method. The sentence compaction is performed by selecting a word set that maximizes a summarization score consisting of the amount of information and the confidence measure of each word, the linguistic likelihood of word strings, and the word concatenation probability. The selected words are concatenated to create a summary. Effectiveness of the proposed method was confirmed by summarizing spontaneous presentations.
Key words Spontaneous speech, presentation speech, automatic speech summarization, sentence extraction, word extraction

1. はじめに

近年の大語彙連続音声認識技術の進展とともに，テキストの読み上げ音声や，アナウンサーが発声したニュース音声などに対しては高精度で認識を行えるようになってきた．しかし，講演音声のような話し言葉の認識においては，60%から70%の認識率となってしまう．一方で，講演・講義などの文字起こしの自動化や，それらのアーカイブへのインデックスの付与，音声を用いたコンピューター対話システムなど，話し言葉音声認識のIT分野への応用に対する要望が高まってきている[1]．

話し言葉の音声認識出力結果には，書き言葉とは異なり，話し言葉特有の言い直し，言い淀み，間投詞などの冗長な表現や，認識誤りの単語も多く含まれている．そこで，このような話し言葉に対しては，音声認識システムにより一字一句残さず文字化するよりも，音声自動要約により不要な単語や文を削除し，話し手が伝えようとした内容・意図を抽出するということが求められる．

自然言語処理の分野でテキストの自動要約に関する研究は既にいくつか行われているが[2]，これらの手法をそのまま話し言葉の音声自動要約に用いても，あまりよい結果は得られない．これは先に述べたように，話し言葉にはどうしてもある程度の認識誤りや不確実性，冗長性が避けられず，書き言葉の要約とは本質的に異なるからである．

音声自動要約に関する研究は近年始まったばかりであり，我々の研究室でも検討を進めてきた．これまで，要約の適正度を示す要約スコアを最大にする単語列を，複数の認識文の中から目的の要約率に合わせて抜き出すという音声自動要約手法を提案しており[3][4]，日本語や英語のニュース音声などに対して有効性が確認されている．しかし，この音声自動要約手法では，認識率の高い文も低い文も，すべての文を一様に扱い，単語抽出によって文を圧縮するという要約を行っている．そのため，ニュース音声に比べて認識誤りや言い誤りなどを多く含む話し言葉にこの手法を用いると，要約率の小さいときに，文として不完全なものや不自然なものを生じやすく，よい要約文が得られない．

そこで本稿では，重要文抽出手法をこれまでの単語抽出による要約手法に組み込んだ，2段階の要約手法を提案する．この手法は，あらかじめ音声認識結果から認識率の低い文，理解困難な文を除き，重要な文を抽出しておく．次に，抽出された文に対して，従来の単語抽出による文の圧縮を行う．このようにすることで，理解困難な要約文の生成を防ぎつつ，できるだけ自然な文となる要約の生成を行う．

本稿の前半では，2段階の要約手法について説明し，後半で，この手法を用いて実際に講演音声の音声自動要約を行った結果とその検討を行う．

2. 音声自動要約手法

本研究で構築した，重要文抽出と単語抽出による要約を組み合わせた音声自動要約システムの構成を図1に示す．

話し言葉音声を入力とし，大語彙連続音声認識の出力として得られる認識文から，はじめに2.1節で説明する各文の要約スコアを計算しておく．次に，認識文からフィラー単語を削除し，残された文の単語数からユーザーによって与えられた目的要約率と重要文抽出による要約の割合を算出し，認識文を重要文抽出，単語抽出の順で，それぞれの要約スコアをもとに要約する．

2.1 重要文抽出

重要文抽出は認識された各文に付けられた，以下のスコアを用いて行う．1文が N 個の単語からなる認識単語列 $W = w_1, w_2, ..., w_N$ のとき，重要文抽出に用いる要約スコア $S(W)$ を以下のように定義する．

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

L, I, C はそれぞれ言語スコア，重要度スコア，信頼度スコアであり，従来までの単語抽出による要約手法[5]を参考に設定した．λ_I, 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．

以下，個々のスコアについて詳しく説明する．

言語スコア

言語スコア $L(w_i)$ は単語連鎖の適正度を表すスコアであり，以下の式で表される．

$$L(w_i) = \log P(w_i | \ldots w_{i-1}) \qquad (2)$$

$P(w_i|\ldots w_{i-1})$ の計算には1.5M形態素からなる，話し言葉コーパスの講演書き起こしテキストから作成した単語trigramを用いた．このスコアは認識誤りによって言語的に不自然な単語連鎖が生じた箇所でより小さい値となり，ペナルティーとして働く．

重要度スコア

重要度スコア $I(w_i)$ は原文の中での相対的な重要度を表すスコアであり，単語の出現頻度に基づく情報量から求める．スコアは，名詞，動詞，形容詞，未知語の内容語に付与され，式(3)で定義される[6]．

$$I(w_i) = f_i \log \frac{F_A}{F_i} \qquad (3)$$

w_i : 音声認識結果に含まれる内容語
f_i : 音声認識結果中の内容語 w_i の出現頻度
F_i : 大規模コーパス中での内容語 w_i の出現頻度
F_A : 大規模コーパス中での総内容語数 $(= \sum_i F_i)$

内容語以外の単語については，重要度スコアを0とする．

講演の書き起こし(1.5M形態素)，60講演の予稿集，WWW上の講演録(2.1M形態素)，NHKのニュース原稿(22M形態素)，毎日新聞(87M形態素)および「音声情報処理」(51k形態素)のテキストコーパスを用い，出現した全約120k種類の単語の各々の出現頻度を求めた．このスコアはキーワードとなる重要な単語に対して大きい値となり，逆に，認識誤りの単語のように原文の内容と関係のない単語では小さい値となる．

信頼度スコア

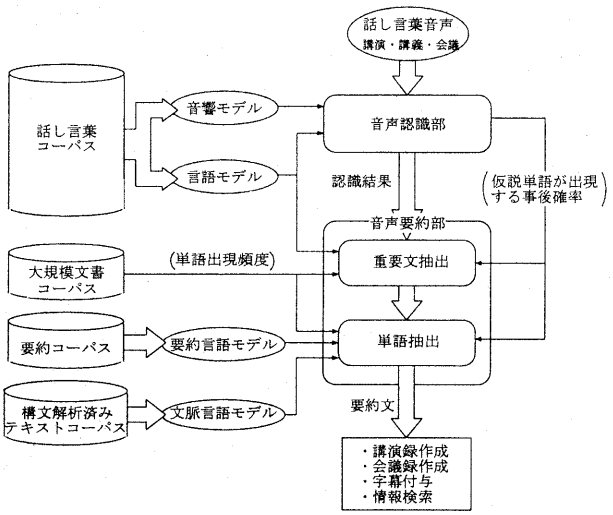

図 1　重要文抽出と文圧縮による音声自動要約システム.

Fig. 1　Automatic speech summarization system using important sentence extraction and sentence compaction.

信頼度スコア $C(w_i)$ は，音声認識結果の音響的，言語的な信頼度を表すスコアである．デコーダから出力された単語グラフに付与された，単語仮説 w_i が出現する事後確率の対数値を各単語に与える．このスコアは音響尤度および言語尤度から計算され，音響的，言語的に信頼度の低い単語には小さい値が付けられる [3].

2.2　単語抽出部

単語抽出部では，前処理として話し言葉と書き言葉の違いを吸収するため，文章を論説調の表現に変換した後，従来の単語抽出による要約手法 [5] を用いて最終的に出力される要約文を作成する．

この単語抽出部で用いる要約スコアには，2.1 節で説明した言語スコア (L)，単語重要度スコア (I)，信頼度スコア (C) と，要約文内の単語連鎖を原文の係り受け関係に基づき制約する単語間遷移スコア (T) を用いる．

単語抽出に用いる要約スコア $S(V)$ は，重要文抽出部で用いたスコアと異なり，単語抽出後の部分単語列 $V = v_1, v_2, ..., v_M$ $(M < N)$ に対して定義される．

$$S(V) = \sum_{m=1}^{M} \{ L(v_m) + \lambda_I I(v_m) + \lambda_C C(v_m) + \lambda_T T(v_m) \} \qquad (4)$$

λ_I，λ_C，λ_T は，各スコアのバランスをとるための重み係数である．この要約スコアが最大となるような単語の組合せを，複数の発話文から 2 段 DP により抽出する．この手法は，第一段階として，可能な全ての要約率で各文を要約し，第二段階として，全体が目的の要約率となるよう各文の要約文を組み合わせ，その中から要約スコアが最大となる組み合わせを動的計画法により決定するというものである (図 2).

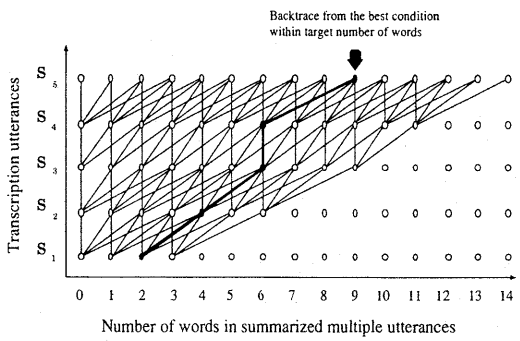

図 2　複数発話の音声要約のための動的計画法の計算領域の例

Fig. 2　An example of DP process for summarization of multiple utterances.

言語スコア $L(v_m)$ の計算に使用する言語モデルは，要約文における単語連鎖をモデル化するものであるが，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．そのため，重要文抽出部で単語 trigram を作成するのに用いた話し言葉コーパスの講演の書き起こし (1.5M 形態素) を論説調の表現に変換したものと，60 講演の予稿集から作成した単語 trigram を用いてスコアを計算した．

単語抽出用の単語重要度スコア，信頼度スコアはともに，重要文抽出用の各スコアの学習に用いたコーパスから求めた．単語間遷移スコアは毎日新聞約 4 万文の構文解析済みの京大テキストコーパスを用いて，構文木制約付きの Inside-Outside アルゴリズムにより，係り受けパラメータの推定を行ったものから求めた．

3. 評価実験

3.1 要約実験条件

日本語話し言葉コーパス (CSJ) 中の男性話者 3 名，

- A01M74 (講演時間 12 分，単語正解精度約 70%，略称 M74)
- A01M35 (講演時間 28 分，単語正解精度約 60%，略称 M35)，
- A05M0031 (講演時間 27 分，単語正解精度約 65%，略称 M31)，

による講演音声を，事前に人手によって意味のある言葉のまとまりごとに切り出して文とし，各文ごとに音声認識を行う．その結果を 70%と 50%の要約率でそれぞれ自動要約を行った．

今回用いた音声認識システムの概略を以下に示す．

特徴抽出

音声データを 16kHz，16bit でディジタル化し，フレーム長 24ms，フレーム周期 10ms で対数パワーと 12 次元のメルケプストラムおよびΔメルケプストラム (計 25 次元) を抽出する．さらに発話毎にケプストラム平均正規化を行う．

音響モデル

話し言葉コーパス中の自動要約対象となる講演者以外の男性話者による 59 時間 (338 講演) の音声データを用いて，16 混合ガウスの不特定話者音素文脈依存 HMM(3000 状態) の音響モデルを作成した．

言語モデル

単語 bigram，trigram を用いる．音響モデルを作成した際に用いた音声データの書き起こし文を，形態素解析システム JTAG により形態素に分解し，約 1.5M 形態素を用いて語彙 20k の言語モデルの学習を行った．ただし，「単語＋読み＋品詞」を形態素の単位とした．

デコーダ

単語グラフを中間表現とする 2 パスデコーダを用いる．第一パスでは 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．

従来手法に，重要文抽出法を組み込んだことによる要約精度の向上効果をみるため，重要文抽出手法を用いないもの (NOS)，言語スコア (L)，重要度スコア (I)，信頼度スコア (C) をそれぞれ単独で用いた重要文抽出手法を取り入れたもの，それぞれのスコアを組み合わせた重要文抽出手法 (L_I，I_C，C_L)，さらにこれら全てのスコアを組み合わせた重要文抽出手法を取り入れたもの (L_I_C) について，全 8 種類の手法による自動要約文をそれぞれの講演音声に対して生成した．ただし，重要度スコア，信頼度スコアの重み係数 λ_I，λ_C，および，重要文抽出手法と単語抽出手法での要約の割合は実験的に求めた最適値を用いた．

参考のため，自動要約文と等しい要約率で単語をランダムに抽出した要約文 (RDM) に対しての評価も行った．

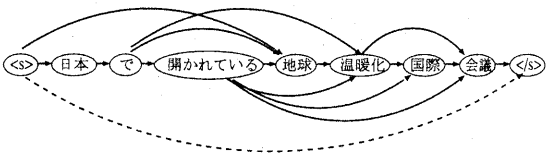

図 3 正解要約文単語ネットワーク

Fig. 3 Word network representing correct summarization.

3.2 実験方法

作成した自動要約文を評価するため，被験者 9 人にそれぞれの講演音声の正解書き起こし文から，単語抽出により要約を作成してもらった．しかし，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ため，正解と考えられる要約文を全て網羅することは困難である．そこで，被験者の作成した要約文をもとに，正解要約文の単語連鎖をまとめた正解要約文単語ネットワーク (図 3) を作成する．この正解要約文単語ネットワークは，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していると考えられる．自動生成した要約文に最も近い単語列をネットワーク中から抽出し，それを正解とすることで，自動生成要約文を一元的に評価できる [3]．要約正解精度は，抽出された正解単語列と自動生成要約文との一致度として，以下の式で定義される値を用いる．

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

Sum_acc : 要約正解精度
Sub : 置換誤り
Ins : 挿入誤り
Del : 削除誤り
Len : 正解単語列の単語数

3.3 実験結果

3.3.1 要約正解精度

要約率 50%，70%のときの実験結果を図 4，5 に示す．全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，従来の手法，本手法とも要約精度が高くなり，手法の有効性が示された．また，要約率 50%，70%のどちらの要約文においても重要文抽出手法を用いない単語抽出のみの要約手法の場合より，両手法を組み合わせた 2 段階の要約手法の要約精度が高いことが示された．

重要文抽出に用いるスコアを単独に用いた場合を見てみると，重要度スコア (I) を用いた重要文抽出手法による要約精度の改善が最も大きく，信頼度スコア (C)，言語スコア (L) の効果がそれに次いでいる．2 つのスコアの組み合わせによる重要文抽出手法 (L_I，I_C，C_L)，さらに全てのスコアを組み合わせた重要文抽出手法 (L_I_C) は，スコアを単独に用いたときよりも要約精度が向上している．

重要文抽出を組み合わせる効果は，要約率がより小さい (圧縮の度合いがより大きい)50%の場合の方が，70%の場合よりも大きく，50%の場合は，3 講演の全てで重要文抽出を組み合わせる効果が見られるが，70%の場合は，効果のある講演とほと

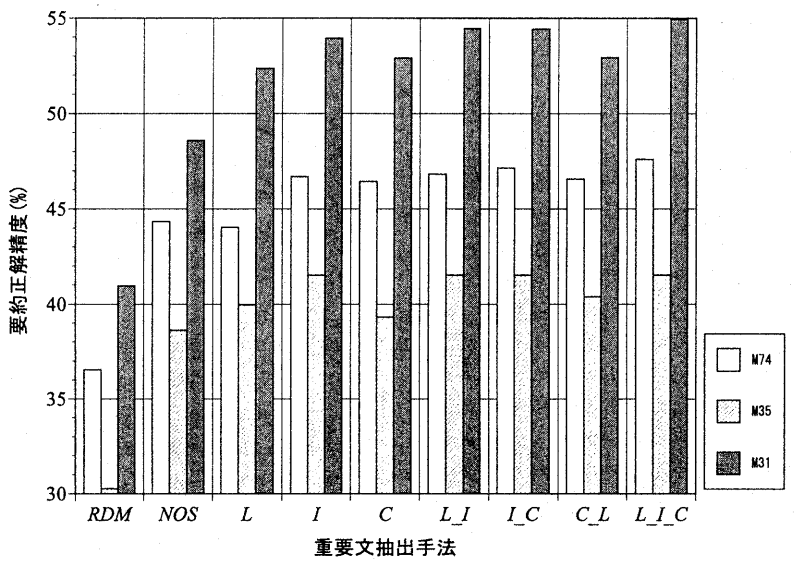

図 4　要約率 50%の要約精度
Fig. 4　Summarization at 50% summarization ratio.

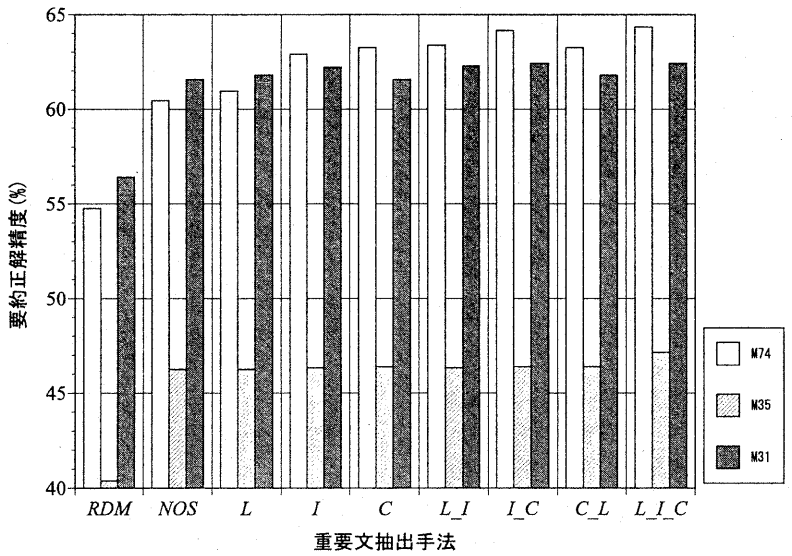

図 5　要約率 70%の要約精度
Fig. 5　Summarization at 70% summarization ratio.

んど効果の見られない講演がある．50%に要約する場合，最も効果の大きい M31 の講演では，重要度スコアによる重要文抽出によって約 5%の要約精度の改善が得られ，全てのスコアを用いた重要文抽出によって約 6%の改善が得られている．70%に要約する場合は，M74 では，重要度スコアによる重要文抽出で約 2%，全てのスコアを用いた重要文抽出によって約 4%の改善が得られている．

要約率 70%の場合に，重要文抽出を組み合わせる効果が比較的小さい M35 と M31 は，フィラーや言い直し，言い誤りなどの冗長な表現が比較的多い特徴がある．要約率が高い場合は，これらに注目して単語単位で削除するのが最も効果的で，文単位の処理による効果は小さいことを示している．

3.3.2　重要文抽出による要約の割合と要約精度

音声認識結果から，まずフィラーを削除し，その後に行う要約処理の内，重要文抽出によって要約する (圧縮する) 割合を種々に変えて，要約精度との関係を関係を調べた．実験結果を

図6と7に示す．これらの図の横軸に示されている「重要文抽出による要約の割合」とは，重要文抽出によって削除した単語数を，目的要約率の要約文作成のために削除した単語数（ただしフィラーは除かれている）で割ったもの，すなわち全体の要約の中で重要文抽出で要約された割合である．

この結果から，重要文抽出による要約の割合の最適値は，講演によって異なるが，要約率が小さくなる（圧縮率が大きくなる）ほど，その割合を大きくした方がよいことが分かる．M74とM31では，その最適値は，要約率70%のとき約0.5(重要文抽出と文圧縮による要約が1:1)，要約率50%のとき約0.7(同じく2:1)となっている．M35では，それぞれ(1:5)と(1:2)となっている．横軸が0と1の場合の結果の比較から，フィラー，言い直しなどが特に多いM35では，重要文抽出だけでは不十分で，むしろ単語抽出による文圧縮が効果的であることが分かる．

重要文抽出と単語抽出による要約の割合をそれぞれどの程度にするかは，要約率とフィラー，言い直しなどの頻度に応じて決定する必要がある．

4. ま と め

本稿では，重要文抽出手法を組み込んだ2段階の音声自動要約手法を提案した．提案法は，従来の単語抽出による要約文作成手法の前段階に重要文抽出手法を組み込み，日本語の文章として不完全な文，理解困難な文をあらかじめ除いておき，残された文に対して単語抽出による要約を行うことで，要約文の高精度化をはかるものである．

講演音声の自動要約を行い，正解要約文単語ネットワークにより要約精度を評価したところ，この手法の有効性が示された．また，複数のスコアを組み合わせることにより，要約精度をより改善できることが示された．

さらに，本手法において，要約率が小さいときは重要文抽出の割合を大きく，要約率が大きいときは重要文抽出の割合を小さくすることで，より要約精度の高い要約文が得られることが示された．

今後の課題として，より要約精度の高い要約スコアの検討，文の認定基準と要約精度の関係についての検討，フィラー，言い直しの頻度に応じた最適な重要文抽出の割合の決定法の検討，被験者の主観評価実験による種々の手法の比較検討などを行っていきたい．

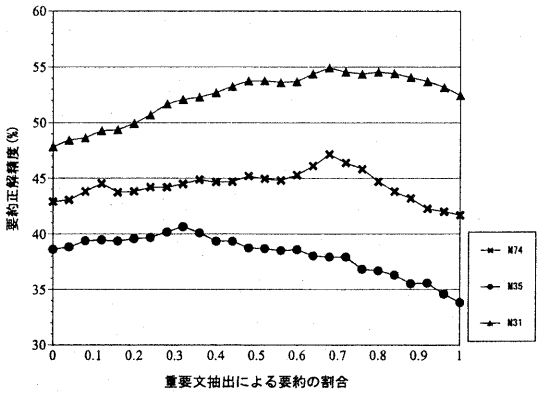

図6 50%要約文における重要文抽出による要約の割合と要約精度
Fig. 6 Summarization accuracy at 50% summarization ratio as a function of the ratio of compression by sentence extraction in the total summarization ratio.

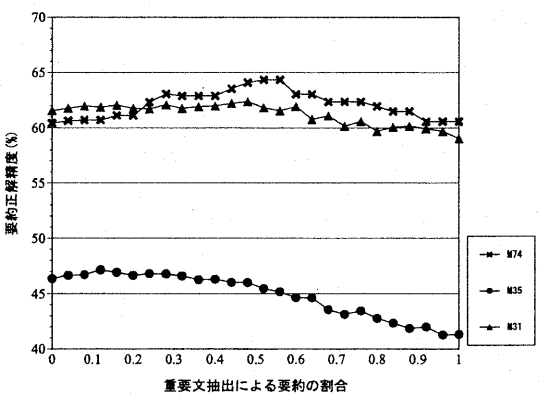

図7 70%要約文における重要文抽出による要約の割合と要約精度
Fig. 7 Summarization accuracy at 70% summarization ratio as a function of the ratio of compression by sentence extraction in the total summarization ratio.

文　献

[1] S. Furui, K. Iwano, C. Hori, T. Shinozaki, Y. Saito and S. Tamura, "Ubiquitous speech processing," Proc. ICASSP2001, Salt Lake City, U.S.A., vol.1, pp.13-16 (2001-5).

[2] 奥村学，難波英嗣，"テキスト自動要約に関する研究動向，" 自然言語処理，vol. 6, no.6, pp.1-26 (1999-4).

[3] 堀智織，古井貞熙，"単語抽出による音声自動要約文生成方とその評価，" 電子情報学会論文誌 D-II，Vol. J85-D-II, No.2, pp.200-209 (2002-2).

[4] 堀智織，古井貞熙，"音声自動要約手法の英語ニュース音声の適用，" 日本音響学会 2002 年春季講演論文集，1-5-12，pp.23-24 (2002-3).

[5] 堀智織，古井貞熙：“講演録作成を目的とした講演音声自動要約，" 日本音響学会 2001 年秋季講演論文集，2-1-10，pp.67-68 (2001-10).

[6] 岩崎淳，古井貞熙：“ニュース音声からの話題抽出法の検討，" 日本音響学会 2001 年秋季講演論文集，1-1-14，pp.27-28 (1998-10).