論文 / 著書情報
Article / Book Information

論題(和文)	重要文抽出と文圧縮による英語音声の自動要約
Title	
著者(和文)	Heie Matthias, 岩野 公司, 古井 貞熙, 堀 智織
Author	Matthias Hietland Heie, Koji Iwano, SADAOKI FURUI
出典(和文)	第3回話し言葉の科学と工学ワークショップ 講演予稿集, Vol. , No. , pp. 105-110
Journal/Book name	, Vol. , No. , pp. 105-110
発行日 / Issue date	2004, 2

重要文抽出と文圧縮による英語音声の自動要約

ヘイエ　マティアス† 　岩野　公司† 　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科学基礎研究所 知能情報研究部
〒 619–0237 京都府相楽郡精華町光台 2–4
E-mail: †{heie,iwano,furui}@furui.cs.titech.ac.jp, ††chiori@cslab.kecl.ntt.co.jp

あらまし　重要文抽出と単語抽出の 2 段階処理に基づく音声自動要約手法を英語ニュース音声に適用した結果について報告する．この手法では，第一段階で予め音声認識結果から認識率の低い文，理解が困難と判断される文を除き，第二段階で，単語抽出による要約により文圧縮を行う．単語抽出による要約に重要文抽出を組み込むことで，単語を単位とした自由度の高い要約文生成を実現しつつ，出現位置の離れた単語の連結による不自然な要約文の生成を抑制することが可能となる．CNN ニュースの 5 つの英語ニュース音声を対象とした要約実験の結果，単語抽出より重要文抽出の効果が特に大きいことが分かった．40% 要約で約 75%, 70% 要約で約 81% の要約正解精度が得られた．
キーワード　音声自動要約，重要文抽出，文圧縮，英語ニュース音声

Automatic speech summarization based on sentence extraction and compaction for English speech

Matthias HEIE†, Koji IWANO†, Sadaoki FURUI†, and Chiori HORI††

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

Abstract　This paper reports experimental results for applying our speech summarization method based on two-stage processing to English broadcast news speech. In the first stage, unreadable and/or less confident sentences are removed and important sentences are extracted from speech recognition results. In the second stage, sentence compaction is performed for the extracted sentences using a word-based extraction method. The two stage scheme can flexibly choose a set of important words from recognition results to be concatenated for creating summary, and can reduce generation of unnatural sentences due to the connection of unrelated words. Experimental results using five English news speech articles excerpted from CNN news show that sentence extraction is more effective than sentence compaction for this task. Summarization accuracies of approximately 75% at 40% summarization ratio and 81% at 70% summarization ratio were obtained.
Key words　automatic speech summarization, sentence extraction, sentence compaction, English broadcast news speech

1. はじめに

近年，音声認識性能は格段に向上し，ニュース音声のような,「書き言葉」の読み上げについては，非常に高い精度で認識を行うことが出来るようになった．しかし，講演や対話など，自発性を伴う「話し言葉」音声に対しては，発話の内容や発声のスタイルが多様であることや，発音変形が多いことなどから，音響・言語特徴の統計モデル化が困難となり，十分な認識精度が得られていない．

そこで，これらの問題を解決するために,「日本語話し言葉工学」プロジェクトによって大規模なコーパス「日本語話し言葉コーパス（CSJ)」が整備された．このコーパスを用いて，講演音声認識に対するモデル改良の研究が進められ，現在，単語正解精度はおおよそ 80% に達したが，未だ十分な認識精度には至っていない．このように，話し言葉の音声認識結果には，認識誤りが多く含まれる．また，間投詞や冗長な表現なども多く含まれていることから，人間が認識結果をそのまま読んでも内容を即座に理解することは難しい．そこで，話し言葉音声に対して，話し手が伝えようとした内容や意図を抽出しつつ，不要な部分を削除し，読み手にコンパクトで分かりやすい文を提示するための「音声自動要約」技術が強く要求されている．

我々はこれまでに，重要文抽出と文圧縮の 2 段階処理に基づく，音声自動要約手法を提案し，日本語話し言葉コーパス（CSJ）の音声を用いた実験により，提案手法の有効性を確認している [1]．この手法では，第一段階で予め音声認識結果から認識率の低い文，理解が困難と判断される文を除き，第二段階で，単語抽出による要約 [2] により文圧縮を行う．単語抽出による要約に重要文抽出を組み込むことで，単語を単位とした自由度の高い要約文生成を実現しつつ，出現位置の離れた単語の連結による不自然な要約文の生成を抑制することが可能となる．

本論文では，この音声自動要約手法を英語ニュース音声に適用した結果について報告する．我々は既に，文圧縮のみに基づく要約手法を，英語音声に適用した結果について報告を行っている [3]．そこで，ここでは，重要文抽出手法の組み込みによる効果について論ずる．

2. 音声自動要約手法

重要文抽出と文圧縮による音声自動要約システムを図 1 に示す.

まず，ユーザーは要約率を設定する．要約率は，原文の単語数に対する要約文の単語数の割合として定義される．また，重要文抽出と単語抽出，それぞれでどれだけの要約を行うか，割合を設定する．システムは，音声認識を行ったのち，認識文からフィラー単語を削除する．残された文について重要文抽出，単語抽出の順で，それぞれの要約スコアをもとに要約処理を行う.

2.1 重要文抽出

重要文抽出では，入力文ごとに以下で定義される要約スコアを求め，スコアの上位となる文を設定した要約率になるまで選択する.

1 文が N 個の単語からなる認識単語列 $W = w_1, w_2, ..., w_N$ に対する，要約スコア $S_s(W)$ は以下のように定義される.

$$S_s(W) = \frac{1}{N}\sum_{n=1}^{N}\{L_s(w_n) + \lambda_{I_s} I_s(w_n) + \lambda_{C_s} C_s(w_n)\} \tag{1}$$

L_s, I_s, C_s 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，$\lambda_{I_s}, \lambda_{C_s}$ 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 $L_s(w_i)$ は各文の単語連鎖の適正度を表すスコアであり，以下のように単語 trigram を用いて算出する.

$$L_s(w_i) = \log P(w_i | w_{i-2}, w_{i-1}) \tag{2}$$

このスコアは認識誤りによって生じる文中の言語的に不自然な単語連鎖に対し，ペナルティーを与える働きがある.

重要度スコア

重要度スコア $I_s(w_i)$ は原文の中での相対的な文の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算出される.

$$I_s(w_i) = f_i \log \frac{F_A}{F_i} \tag{3}$$

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 $(= \sum_i F_i)$

スコアは内容語のみに付与され，それ以外の単語についてはスコアを 0 と定義した．なお，本研究では，名詞と動詞を内容語とした.

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が付けられる.

信頼度スコア

信頼度スコア $C_s(w_i)$ は，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で定義される．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.

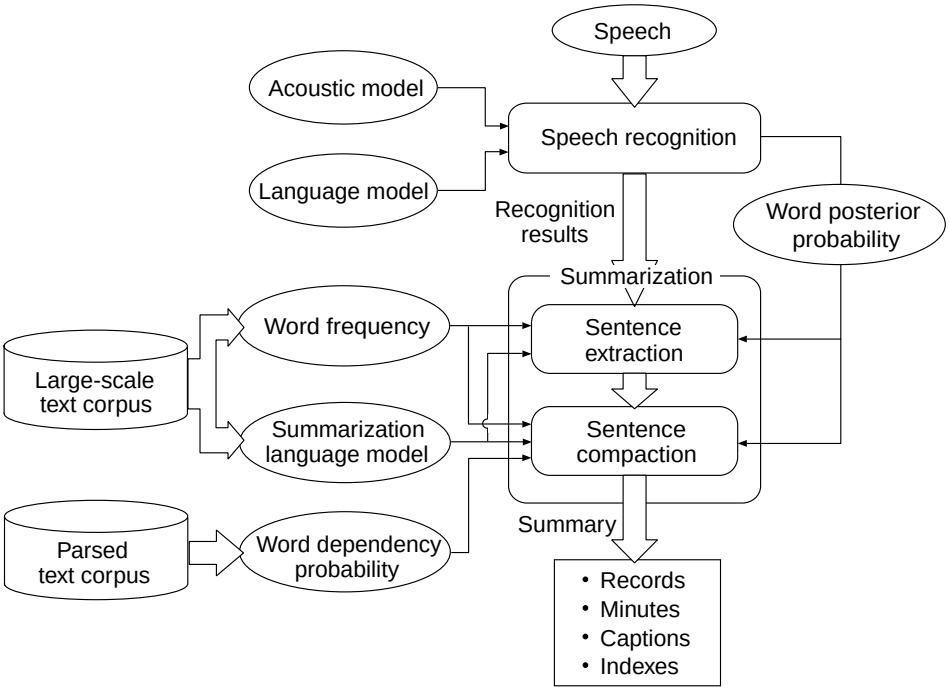

図 1　重要文抽出と文圧縮による音声自動要約システム

Fig. 1　Automatic speech summarization based on sentence extraction and compaction.

2.2 単語抽出

単語抽出は，文献 [3] と同様に行った．

要約スコアとしては，重要文抽出と同様の「言語スコア (L_w)」「重要度スコア (I_w)」「信頼度スコア (C_w)」の他に，要約文の単語間の係り受け構造を考慮した「単語間遷移スコア (T_w)」を用いる．そのため，構文解析済みのテキストコーパスから単語単位の係り受け SCFG を学習しておく．

単語抽出に用いる要約スコア $S_w(V)$ は，重要文抽出と異なり，単語抽出後の部分単語列 $V = v_1, v_1, \cdots, v_M \quad (M < N)$ に対して以下のように定義される．

$$S_w(V) = \sum_{m=1}^{M} \{L_w(v_m) + \lambda_{I_w} I_w(v_m) + \lambda_{C_w} C_w(v_m) + \lambda_{T_w} T_w(v_m)\} \quad (4)$$

$\lambda_{I_w}, \lambda_{C_w}, \lambda_{T_w}$ 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 $S_w(V)$ が最大となるような部分単語列を 2 段 DP 法に基づき決定することで，単語抽出による要約処理を行う．

3. 評価実験

3.1 実験条件

NIST 主催の Topic Detection and Tracking (TDT) タスク中の，5 つの CNN ニュース（128 発話文）を要約対象とし，これらを，要約率 40%, 70% で要約した．単語には Brill tagger [4] を用いて，品詞情報が付加されており，評価時にも品詞情報を考慮した．

評価には，要約正解精度 [2] を用いる．そのため，予め複数の被験者が作成した正解要約文から，単語間の連鎖をまとめた単語ネットワークを作成しておく．評価の際には，自動要約文に一番近い単語列をネットワークから抽出し，その単語列を正解文とみなしたときの要約文の単語正解精度を要約正解精度とする．本実験では，英語を母国語とする 17 名の被験者が作成した要約正解文から，単語ネットワークを生成し，要約正解精度を求めた．

文献 [3] では，同じ評価データ・実験条件における，単語抽出のみによる要約正解精度を報告している．単語抽出における 4 つの要約スコアを，最適な重みで融合して用いたときに最も高い性能を示し，その時の要約正解精度は 40% 要約で 54.1%, 70% 要約で 71.1% であった．本実験では重要文抽出を組み込んだ効果についての評価を行うため，単語抽出部はこの最適化された条件で固定しておく．

3.2 音声認識部

音声認識には JRTk (JANUS Speech Recognition Toolkit) [5] を用いた．認識システムの条件を以下に示す．

特徴量抽出

16kHz, 16bit でデジタル化した音声データを，13 次元の MFCC に変換し，声道長正規化 (VTLN) および，クラスタごとのケプストラム正規化を行う．次に，7 フレー

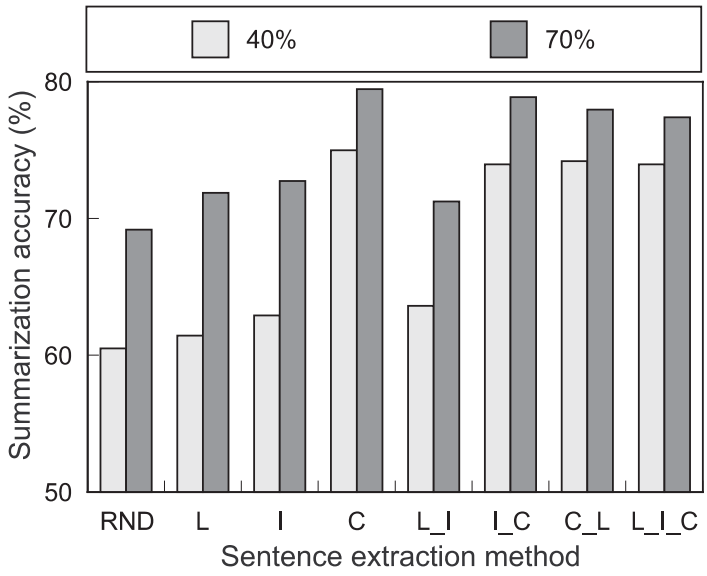

図 2　要約スコアごとの重要文抽出による要約正解精度

Fig. 2　Summarization accuracies by sentence extraction for various summarization conditions.

ムを 1 セグメントとして各セグメントを LDA (Linear Discriminant Analysis) を用いて 42 次元に圧縮し，特徴量とする．

音響・言語モデル

使用した音響・言語モデルは，文献 [3] と同様である．音響モデルは，状態数 6k の quinphone HMM であり．各混合分布は 2k 個のガウス分布コードブックを重み付けし共有化している．合計のガウス分布数は 105k である．音響モデルの学習データには, Wall Street Journal, English Spontaneous Scheduling Task, Broadcast News, Crossfire and Newshour TV news show を用いている．

言語モデルには, Broadcast News コーパスから学習した語彙サイズ 40k の単語 trigram を用いている．

デコーダ

JRTk を用いて，単語グラフを中間表現とする以下の 3 パスの探索を行う．第 1 パスでは，triphone と bigram を用いて木構造辞書に基づく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．第 2 パスでは，単語グラフ上の単語からフラットな辞書を構成し，quinphone と trigram を用いてビームサーチを行い単語グラフを再構成する．第 3 パスでは，単語グラフを最小化した後で trigram を用いてリスコアを行い，最終的な認識結果を得る．

なお，評価データに対する単語正解精度は 78.4% であった．

3.3 要 約 部

言語スコア算出のための trigram，重要度スコア算出のための単語頻度情報は，Penn Treebank コーパス [6] に含まれる Wall Street Journal と Brown コーパス（約 175k 文，4.7M 単語）から作成した．単語抽出に用いる SDFG は，Brown コーパス中の約 11k 文を使って学習した．

3.4 実験結果

3.4.1 重要文抽出の最適化

まず，重要文抽出における最適な要約スコアの組み合わせを検証した．要約スコアとして，言語スコア (L_s)，重要度スコア (I_s)，信頼度スコア (C_s) のみを使用した場合，それぞれのスコアを組み合わせて使用した場合 (L_I, I_C, C_L)，全てを組み合わせて使用した場合 (L_I_C) について実験を行った．重み係数 $\lambda_{I_s}, \lambda_{C_s}$ は，ニュースごとに，他の 4 つのニュースを用いた実験結果が最良となるように最適化を行った．また比較用として，ランダムに重要文抽出を行った場合の結果を RND として示した．

40%, 70% 要約における重要文抽出のみでの要約正解精度を図 2 に示す．ランダム選択と比較して，全ての要約スコア条件で要約正解精度の向上が確認できる．特に，信頼度スコア C_s が有効であることがわかる．言語スコア L_s，重要度スコア I_s を単独で用いた場合は，ランダム選択に比べ若干効果が得られているものの，信頼度スコア C_s と組み合わせた時には，その効果が見られなかった．

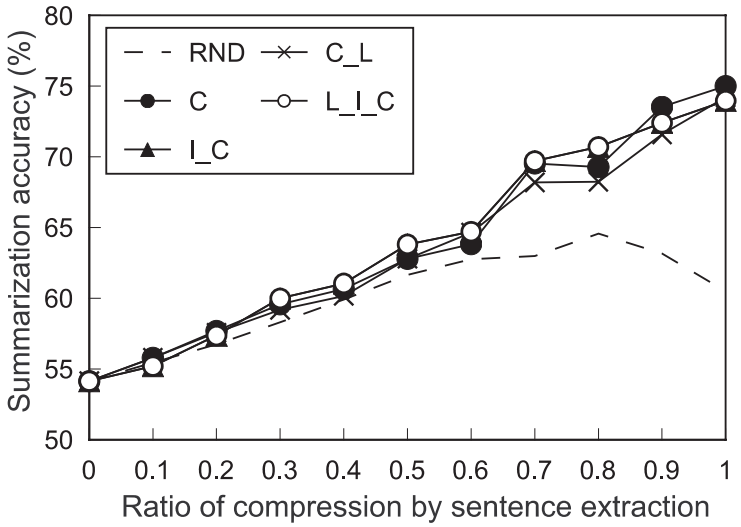

図 3　40% 要約における重要文抽出による要約の割合と要約正解精度

Fig. 3　Summarization accuracy at 40% summarization ratio as a function of the ratio of compression by sentence extraction in the total summarization ratio.

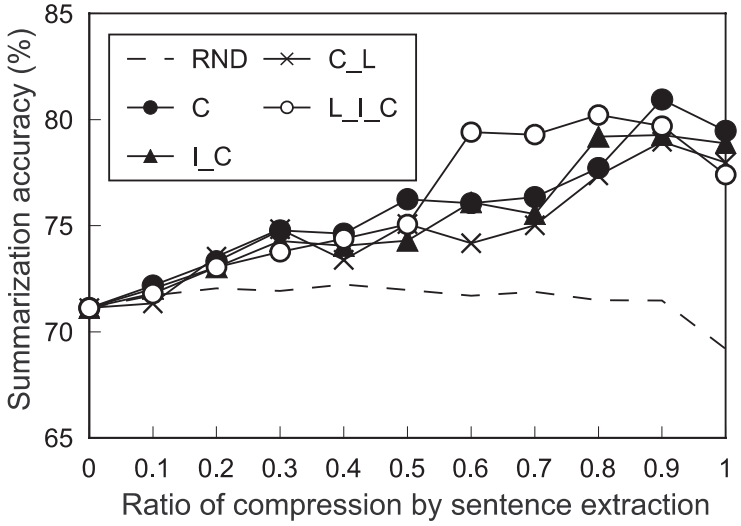

図 4　70% 要約における重要文抽出による要約の割合と要約正解精度

Fig. 4　Summarization accuracy at 70% summarization ratio as a function of the ratio of compression by sentence extraction in the total summarization ratio.

3. 4. 2　重要文抽出と単語抽出による要約結果

そこで，信頼度スコア C_s を含む種々の要約スコアを利用した場合について，重要文抽出・単語抽出を組み合わせた手法の要約性能の評価を行う．

重要文抽出による要約の割合を 0 ～ 1 まで 0.1 刻みで変化させた時の 40%, 70% 要約時の要約正解精度の変化をそれぞれ図 3, 4 に示す．横軸が重要文抽出による要約の割合であり，0 が単語抽出のみを行った場合，1 が重要文抽出のみを行った場合に相当する．重み係数には，重要文抽出のみによる実験において最適化された値をそのまま用いている．

40%, 70% 要約，どちらの場合においても，スコアの組み合わせの違いによる，大きな傾向の違いは見られなかった．最も良い結果は，信頼度スコア C_s のみを用いた時に観測された．

40% 要約では，重要文抽出による要約の割合に対して単調に要約正解精度が向上している様子がわかる．重要文抽出のみを行った場合に最も性能がよく，単語抽出のみによる結果と比較すると絶対値で 20.9%，要約正解精度が向上し，最高の要約正解精度は 75.0% となった．しかし，両手法を組み合わせた効果は得られなかった．

70% 要約では，重要文抽出による要約の割合が 0.8 ～ 0.9 の付近で，重要文抽出と単語抽出の組み合わせ効果が得られ，単語抽出のみの結果から最高で 9.8% の要約正解精度の向上が見られ，要約正解精度は 81.0% に達した．最適な重要文抽出による要約の割合が 1 に非常に近

いことから，70% 要約の場合においても重要文抽出の効果が大きいことがわかる．

4. おわりに

重要文抽出と単語抽出の 2 段階処理に基づく音声要約手法を英語ニュース音声に適用した結果について報告した．重要文抽出を組み込むことによって，単語抽出のみでの要約正解精度が大きく改善されることが分かった．また，70% 要約では，両者を適切に組み合わせることで，最も高い要約性能が得られることが示された．

実験結果からは，英語ニュース音声の要約に対し，特に重要文抽出の効果が大きいことが示されたが，同じ手法を日本語話し言葉に適用した結果 [1] からは，これほど顕著な傾向は見られなかった．これは，言語の違いだけでなく，放送ニュースと話し言葉というスタイルの違いにも起因している可能性がある．今後は，日本語ニュース音声，あるいは英語話し言葉を対象とした要約実験の結果との比較を行い，その背景を明らかにする必要がある．その際，被験者による正解要約文の作成の傾向に，要約対象となる音声の言語間，スタイル間で違いがあるかについても調べる必要もある．

また，各種重みパラメータの最適化手法などについても検討して行く必要がある．

謝　　辞

音声認識装置を提供し，英語の音声認識に貢献してくださった Carnegie Mellon University の Alex Waibel 教授，Rob Malkin 氏，Hua Yu 氏に感謝致します．正解要約文作成に協力してくださった Sheffield 大学の Yoshi Gotoh 氏に感謝致します．

文　　献

[1] 菊池智紀，古井貞熙，堀　智織, “音声自動要約における重要文抽出の利用,” 音講論, vol.1, pp.97-98 (2002-9).

[2] 堀　智織，古井貞熙，“単語抽出による音声自動要約文生成法とその評価,” 信学論 D-II, vol.J85-D-II, no.2, pp.200-209 (2002-2).

[3] 堀　智織，古井貞熙，“英語ニュース音声を対象とした音声自動要約,” 音講論, vol.1, pp.69-70 (2001-10).

[4] http://www.cs.jhu.edu/~brill/

[5] A. Waibel, H. Yu, M. Westphal, H. Soltau, T. Schultz, T. Schaaf, Y. Pan, F. Metze, and M. Bett, “Advances in meeting recognition,” *Proc. HLT 2001*, pp.11-13, San Diego (2001-3).

[6] http://www.cis.upenn.edu/~treebank/