韻律情報を利用した重要文抽出に基づく音声自動要約

Automatic Speech Summarization Based on Extraction of Important Sentences Using Prosodic Information

立命館大学情報理工学部メディア情報学科
College of Information Science and Engineering, Ritsumeikan University

山下 洋一
Yoichi YAMASHITA

＜研究協力者＞
立命館大学大学院理工学研究科
Graduate School of Science and Engineering, Ritsumeikan University

笠原 力弥 Rikiya KASAHARA　三上 貴由 Takayoshi MIKAMI　井上 章 Akira INOUE

This paper describes speech summarization based on the extraction of important sentences using prosodic information. A multiple regressive model using linguistic scores and prosodic parameters predicts the importance score of the sentence. The proposed method is evaluated both on the correlation between the predicted sentence importance and the preference scores by human subjects and on the accuracy of extraction of important sentences. Prosodic information improved the quality of speech summary, and it is more effective when the speech is transcribed by automatic speech recognition because speech recognition errors damage linguistic information. This paper also describes a method of segmenting a spoken document into sentences based on a classification tree technique using prosodic information. Evaluation experiments reveal that pause duration and power information are important to identify the end of sentences.
Key words: speech summarization, prosody, sentence extraction, multiple regression, classification tree

1　研究の目的

情報技術の進歩によって、画像、音声、文字テキスト、さらにそれらを組み合わせたメルチメディア情報が大量に蓄積できるようになっている。情報コンテンツの表現においてマルチメディア化が進むにつれて、音声に代表される言情報を含むコンテンツも増大しており、講演や演説など音声言語が中心的な役割を果たすコンテンツも多い。今後も、文化的／歴史的に意義の大きい講演や演説が多数蓄積されていき、大学での授業や学会講演などの学術的コンテンツのディジタルカーカイブ化も進むものと思われる。蓄積された情報コンテンツのデータベースから欲しいデータを探し出す時、一般に、蓄積されたデータ量が増えれば増えるほど、欲しいデータを探し出すことが難しくなり、検索技術やコンテンツへのアノテーションが重要になってくる。

これまでに、文字テキストに対する自動要約の研究が広く行なわれてきている [1, 2]。近年、連続音声認識の性能が向上したことにより、音声データに対する自動要約の研究も始まっている [4, 5, 6, 7, 8, 9, 10]。音声データは、文字テキストと比べてスキャニング (拾い読み／拾い聞き) が難しく、講演などの音声に対する要約自動生成に対する期待は大きい。探し出した音声データが必要なものかどうかを判断するのに、要約された結果を聴いたり読んだりできれば非常に有用である。また、音声の自動要約に関する技術は、会議の収録音声からの議事録自動作成などへ発展していくことが考えられ、重要な要素技術として位置付けられる。

音声データの自動要約は、図 1(a) に示すように、音声を連続音声認識によって文字テキストに変換し、得られた文字テキストに対してテキスト要約 [1, 2] を行なうことによっても実現できる。しかし、このような処理は音声の持つ言語的な情報のみに注目しており、非言語的な (パラ言語的な) 情報が無視されることになる。音声によるコミュニケーションで

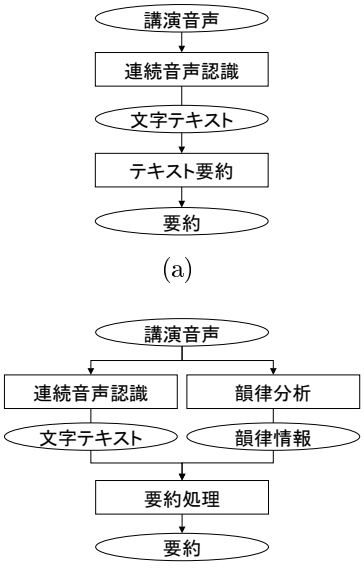


図 1: 講演音声の要約処理過程

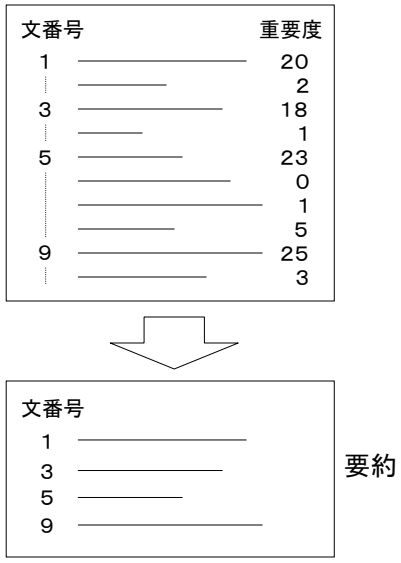


図 2: 重要文抽出による要約

は、意図、感情、強調、微妙なニュアンスなどの非言語的情報が韻律情報 (声の高さ、声の大きさ、発話速度) によって表現されることがよく知られている。音声の要約でも、図 1(b) に示すように、音声波形の持つ韻律情報を言語情報と併せて利用することによって、要約の精度を向上させられる可能性がある。そこで本研究では、講演音声を対象として、言語的な情報に加えて韻律情報を利用して要約を生成する手法を開発することを目的とする。

2 講演音声の要約

人間が文章を要約するときには、まず全文を読んで内容を理解してから重要な箇所を取り出し、それを頭の中で再構成し要約を完成させる。しかし、現在の情報処理研究では、十分な意味理解ができるところまで技術が進んでいないため、計算機が人間と同じような処理過程で要約を作り出すことは難しい。そこで本研究では、講演などの音声データから重要文を抽出する処理を要約と考える。

これまでに行われてきたテキスト自動要約では、図 2 に示すようにあらかじめ決められた文などの単位のうち、重要な部分を抽出することによって抄録する方法が多い [1, 2]。このような処理による講演音声の要約は、以下の過程からなる。

1. 講演音声を文などの単位に分割する。
2. １文毎に重要度を算出する。
3. 重要度の高い文から必要数の文を抜き出す。

重要部分を抜き出すことで要約を生成する場合には抜き出す単位が問題になる。テキスト要約では文が単位として用いられることが多いが、音声要約では文の単位を決定することすら簡単ではない。本報告では、まず、人手で決定した文単位を用いて、文ごとの重要語を算出し重要文を抽出する手法について述べる。さらに、ポーズをもとに分割された発話単位をもとに文境界を自動的に決定する手法についても述べる。

3 韻律情報を用いた文重要度の自動決定

3.1 言語情報

言語情報と韻律情報を組み合わせて文の重要度を算出するために、韻律情報だけでなく、文テキストからの言語情報の獲得が必要となる。これまでの研究から重要な単語（重要語）が多く含まれる文は重要度が高く、出現頻度が中程度の単語は重要語である確率が高いことなどが知られている。これより、文章中での単語の出現頻度を見ることで各文の重要度を算出する事が可能であると言える。ほかにも重要文の検出について言語情報の有用な手がかりとしてこれまでにいくつかの方法が提案されている。要約に用いられる言語情報として、

- 文中の位置（冒頭、段落頭、文章末など）
- 重要語の出現頻度
- 原文の構造を解明

- 文と文のつながり具合
- 手がかり語（「ようするに」「つまり」など）

などが試みられている。このような言語情報の利用に関しては、本研究では公開されているテキスト要約システム Posum[2, 11] を用いて、言語情報を利用することとした。Posum は、テキスト中の単語の重要度や単語間のつながりを利用する基本的な要約エンジンで、テキストを入力とし、各文の重要度を出力することができる。本研究では、この重要度スコアを言語情報として利用し、以下 $LING$ と表記する。

3.2 韻律パラメータ

文の重要度を決定するために、韻律情報を利用することを考える。以下に述べる時間長、パワー (声の大きさ)、基本周波数 (声の高さ) に関するパラメータを文ごとに算出する。

3.2.1 基本周波数

基本周波数に関するパラメータとして、以下に示す文中の最小値 F_{min}, 最大値 F_{max}, レンジ F_{range}, 平均値 F_{avg} の４つのパラメータを文ごとに算出する。

$$\begin{aligned} F_{avg} &= \frac{1}{L}\sum_{i=1}^{L} f_i \\ F_{min} &= \min\{f_1, f_2, \cdots, f_L\} \\ F_{max} &= \max\{f_1, f_2, \cdots, f_L\} \\ F_{range} &= F_{max} - F_{min} \end{aligned}$$

ここで、L はその文のフレーム数、f_i はその文の i 番目のフレームの基本周波数である。基本周波数の算出には、Entropic 社の音声分析ライブラリ ESPS を用いた [12]。

3.2.2 音素時間長

文の j 番目の音素 ph_j の音素長 D_j を次式で正規化し、正規化された音素時間長 d_j を求める。

$$d_j = \frac{D_j - \bar{D}(ph_j)}{\sigma_D(ph_j)} \qquad (1)$$

ここで、$\bar{D}(ph)$ と $\sigma_D(ph)$ はそれぞれ音素 ph の時間長の平均と標準偏差である。各音素の時間長 D_j は、音声認識ツール HTK [13] を用いた強制整列によって求めた。音素時間長に関するパラメータとして、上記の正規化時間長を用いて以下に示す文中の最小値 DUR_{min}、最大値 DUR_{max}、レンジ DUR_{range}、平均値 DUR_{avg} の４つのパラメータを文ごとに算出する。

$$\begin{aligned} DUR_{avg} &= \frac{1}{N}\sum_{j=1}^{N} d_j \\ DUR_{min} &= \min\{d_1, d_2, \cdots, d_N\} \\ DUR_{max} &= \max\{d_1, d_2, \cdots, d_N\} \\ DUR_{range} &= DUR_{max} - DUR_{min} \end{aligned}$$

ここで、N はその文の音素数である。

3.2.3 パワー

文の j 番目の音素の中心 20ms の区間の平均パワー P_j を、音素時間長に対する式 (1) と同様に、音素毎の平均値と標準偏差を用いて正規化した値を p_j とおく。パワーに関するパラメータとして、文中の最小値 POW_{min}、最大値 POW_{max}、レンジ POW_{range}、平均値 POW_{avg} の４つのパラメータを 3.2.2 節の音素時間長と同様に文ごとに算出する。

3.2.4 発話時間長

文の発話時間長は、狭い意味での韻律情報にはあたらないが、予備的な検討から文重要度との関連性が高いことが明らかとなったため、パラメータとして利用する。以下では、これを LEN と表記する。

3.3 音声データ

講演音声データとしては約 10 分のＮＨＫ論説番組「あすを読む」の 5 回分を用いた。表 1 に用いたデータの内容、話者の性別、文数を示す。講演音声の文単位への分割は、人手で行なった。音声認識の性能評価、さらに重要文抽出における認識誤りの影響の分析を行なうために、まず、人手でよる書き起こしテキストを作成した。

3.4 音声認識

言語情報を得るために文ごとに音声認識を行う。音声認識は、CSRC[14]2001 年度版のシステムを用いて行なった。具体的には、デコーダは julius-3.3p3 高精度版 [15]、音響モデルは 64 混合分布、3000 状態の PTM モデル、言語モデルは語彙数 20K の 3-

表 1: 講演音声データ

データ番号	データ 1	データ 2	データ 3	データ 4	データ 5
内容	東海村臨海事故	高齢者パワーをどう活かしていくか	砂浜の再生	原発と老朽化	ヤコブ病訴訟和解へ
話者	男声 A	女性 A	男声 B	男声 C	女性 B
文数	65	68	71	76	71

gram を用いた。5 つの講演音声データに対する平均単語認識精度は 64.6% であった。

3.5 文重要度の決定実験

重要文抽出の性能は、人手で決定した文の重要度に基づいて評価する。人手で作成した書き起こしテキストを用い、以下の要領で重要文抽出の要約実験を行なった。

(1) 番組のビデオの視聴し、概要を理解する。

(2) 書き起こしテキストを見ながら音声を聴収し、書き起こしテキストから重要文／非重要文をそれぞれ 10 文程度抽出する。

被験者数はデータ 1〜5 に対して、それぞれ、14, 18, 13, 14, 15 人である。

この結果から、i 番目の文の重要度 $SI(i)$ は、

$$SI(i) = R(i)_{imp} - R(i)_{unimp} \qquad (2)$$

で求める。ここで $R(i)_{imp}$、$R(i)_{umimp}$ は、それぞれ i 番目の文を重要文として選んだ人の割合、非重要文として選んだ人の割合である。図 3 に要約実験から決定された文重要度の例を示す。

3.6 文重要度と各韻律パラメータとの相関

複数の韻律パラメータを組み合わせて用いる前に、各韻律パラメータと人手による文重要度との相関を調べた。結果を図 4 に示す。ここでは、5 つのデータに対する相関係数の平均値を示している。これより、言語情報 ($LING$) の相関が高いほか、LEN, POW_{avg}, POW_{max}, POW_{range} においても高い相関が見られる。

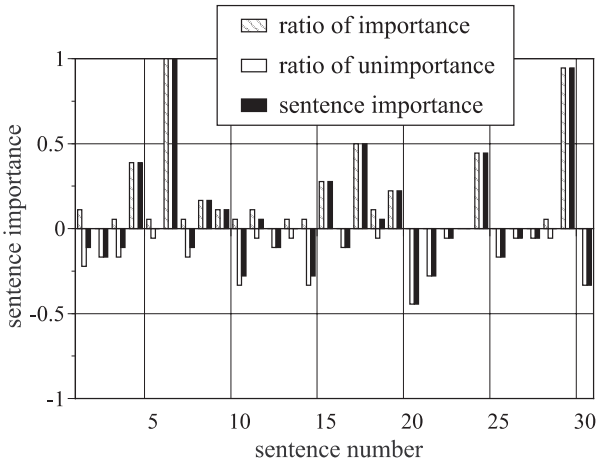


図 3: 要約実験から決定された文重要度の例

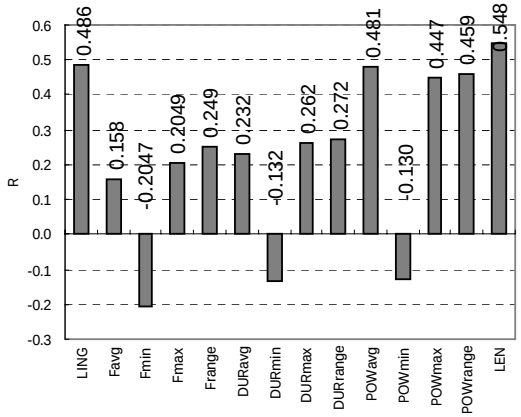


図 4: 重要度と各韻律パラメータの相関係数

3.7 複数の韻律パラメータの組合せによる文重要度の予測

次に、言語情報と複数の韻律パラメータを組み合わせて用い、文の重要度を予測する。予測には重回帰モデルを利用する。言語情報は必ず用いることとし、i 番目の文の重要度を

$$SI(i) \quad = \quad a_0 + a_{LING} \times LING(i)$$

表 2: パラメータの組合せ

パラメータセット	用いるパラメータ
$C0$	$LING$
$C1$	$LING, LEN, F_{range}, DUR_{range}, POW_{avg}$
$C2$	$LING, LEN, F_{range}, F_{min}, DUR_{range}, DUR_{max}, POW_{avg}, POW_{range}$

$$+\sum_{j=1}^{M} a_j \times B(i)_j \qquad (3)$$

で予測する。$LING(i)$ は Posum から出力される言語情報のみによる i 番目の文の重要度スコア、$B(i)_j$ は i 番目の文における j 番目の韻律パラメータ、M は組み合わせる韻律パラメータの数である。学習データを用いて、モデルパラメータ a_0, a_{LING}, a_j を決定することによってモデルが作成される。

パラメータの組合せを考える時、3.2 節で述べた全ての韻律パラメータを用いて重回帰モデルを作成することもできるが、基本周波数に関する $F_{avg}, F_{min}, F_{max}, F_{range}$ など、同じ種類のパラメータ間では相互の相関が高い場合があり、用いるパラメータ数を増やすことが必ずしも精度の高いモデルの作成にはつながらない。そこで、3.7 節での結果をもとに、韻律パラメータの種類 (基本周波数、音素時間長、パワー) ごとに文重要度との相関の高いものから順にパラメータを選択して用いることを考える。用いるパラメータの組合せとして、表 2 に示す 3 つのセットを試みた。$C0$ は、言語情報のみを用いる場合で、これがベースラインの性能を与える。$C1$ は、これに発話時間長を加え、韻律パラメータの種類ごとに、文重要度との相関が最も高いパラメータを一つずつ加えている。$C2$ では、さらに韻律パラメータの種類ごとに、相関が 2 番目に高いパラメータも加えている。

3.7.1 相関係数による評価

各パラメータセットにおける重回帰モデルの重相関係数の値を、書き起こしテキストの作成方法および評価の仕方に分けて図 5 に示す。重相関係数は 0 から 1 の間の値をとり、値が大きいほどモデルによる現象の説明がうまくいっていることを示す。

ここで trans- と CSR- はそれぞれ人手による書き起こしテキストと自動音声認識によるテキストから言語情報スコア $LING$ を算出した場合を示しており、また、-closed、-open はそれぞれクローズド

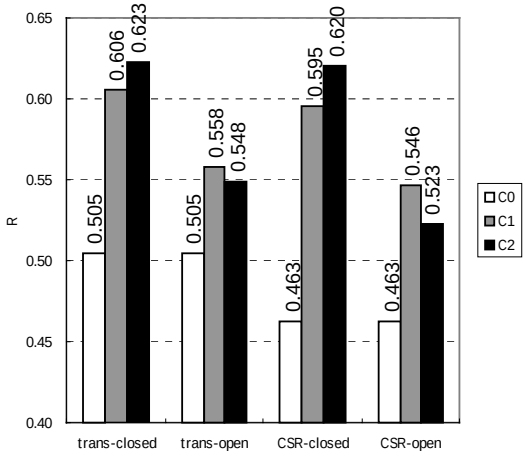


図 5: 重相関係数による評価

評価、オープン評価を示している。オープン評価では、表 1 の 5 つの講演音声データのうち 4 つのデータで重回帰モデルを作成し、残りの一つのデータに適用して評価する処理を 5 回行ない、その平均値を求めた。

図 5 を見ると $C1, C2$ ともベースラインの $C0$ よりも重要関係数が大きくなっている。$C1$ と $C2$ の比較では、オープンな評価において $C2$ の方がやや重相関係数が小さくなっており、必ずしも用いる韻律パラメータが多い方が良いとはかぎらないことを示している。しかし、学習に用いたデータ数がそれほど多くはないため、今後、学習データ量をさらに増やした場合には、$C2$ の方が重相関係数が大きくなることも考えられる。

次に、書き起こしテキストを人手で作成した場合と、連続音声認識の結果を利用した場合を比較する。$C0$、すなわち言語情報だけで重要度を決定する場合には、連続音声認識を用いることによって重相関係数がやや小さくなっている。これは、連続音声認識による認識誤りのために言語情報が劣化したためと考えられる。一方、$C1, C2$ の韻律パラメータを利用して文重要度を決定した場合には、連続音声認識を用いても重相関係数はそれほど変化しておらず、韻律パラメータを利用することによる効果は連続音声認識を使った場合の方が顕著であることがわかる。

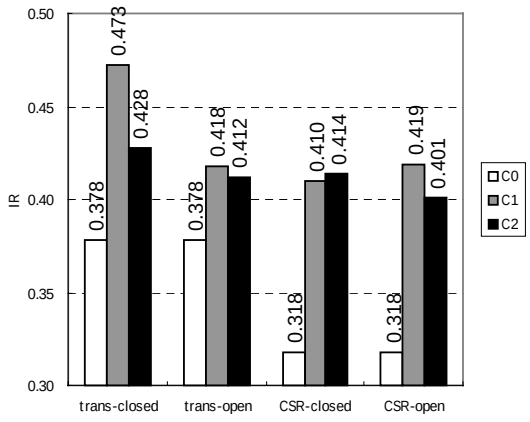


図 6: 重要文認定度による評価

3.7.2 重要文認定度による評価

重要文抽出による要約の精度を評価するために、次式で定義する重要文認定度 IR を新しく導入する。

$$IR = \frac{1}{4}(IR_5 + IR_{10} + IR_{15} + IR_{20}) \quad (4)$$

$$IR_n = \frac{C(n)_{imp} - C(n)_{umimp}}{n} \quad (5)$$

ここで $C(n)_{imp}, C(n)_{umimp}$ は文重要度に基づいて抽出された上位 n 文と、人手による重要度におけるそれぞれ上位 n 文／下位 n 文との一致文数である。IR_n は n 文の重要文抽出を行なうとき、抽出されるべき 1 つの重要文が実際に抽出される「見込み」を表しており、$n = 5, 10, 15, 20$ の時の結果を平均した値を IR としている。

重要文認定度 IR による各パラメータセットに対する評価結果を図 6 に示す。ここで `trans`、`CSR`, `closed`, `open` の意味は、図 5 と同じである。

図 6 を見ると、ベースラインの $C0$ より $C1, C2$ の精度が高くなっており、重要文認定度の尺度においても韻律情報を利用した効果が確認できる。さらに、韻律パラメータを利用することによる効果が連続音声認識を使った場合の方が顕著であることも重相関係数による評価の場合と同様である。

4 韻律情報を用いた文境界の自動決定

4.1 文境界の決定

書き言葉と異なり話し言葉では、文の単位が明示的に与えられない。そのため、重要文の抽出に基づいた音声の自動要約を行うためには、文単位の自動決定が重要な問題となる。

本研究では、300 ミリ秒以上のポーズで区切られた発話区間を基本的な単位 (発話単位) とみなし、文は複数の発話単位で構成されると考える。したがって、発話単位の境界に対して、それが文境界になるかどうかを判断することによって文単位の決定を行う。

発話単位境界が文境界であるかどうかの判断は、言語情報と韻律情報を入力 (説明変数) とする分類 2 進木によって行う。分類 2 進木としては、Salford 社の CART4.0[16, 17] を利用した。

4.2 言語情報

ポーズの直前と直後の品詞を用いる。ポーズ直前の品詞に関しては、文末表現に関わる品詞を考え、「助動詞」「終助詞」「それ以外」の三つのカテゴリに分類した結果を用いる。ポーズ直後の品詞に関しては、「接続詞」、「フィラー・感動詞」「その以外」の三つのカテゴリに分類した結果を用いる。品詞の決定には、形態素解析システム「茶筌」[18] を用いて自動的に決定したものを用いた。

4.3 韻律パラメータ

4.3.1 基本周波数

基本周波数の算出には ESPS を使用し、フレーム長 32ms、フレームシフト 5ms で分析した。基本周波数に関わるパラメータとして、以下の 3 つのパラメータを用いる。

$$F_b = (f_{b1} + f_{b2})/2 \quad (6)$$

$$F_f = (f_{f1} + f_{f2})/2 \quad (7)$$

$$F_{bf} = F_b/F_f \quad (8)$$

ここで、f_{bi} は、ポーズ直前の発話単位における末尾から i 番目の音素の平均基本周波数である。また、f_{fi} は、ポーズ直後の発話単位における先頭から i 番目の音素の平均基本周波数である。F_b, F_f はそれぞれポーズの直前、直後の基本周波数を表わし、F_{bf} はその比を表わしている。

4.3.2 パワー

パワーに関わるパラメータとして、以下の 3 つのパラメータを用いる。

$$P_b = (p_{b1} + p_{b2})/2 \quad (9)$$

$$P_f = (p_{f1} + p_{f2})/2 \quad (10)$$

$$P_{bf} = P_b - P_f \quad (11)$$

表 3: 講演音声データ

データ番号	内容	発話単位数
データ 1	東海村臨海事故	177
データ 3	砂浜の再生	181
データ 6	道路公団の改革	163

ここで、p_{bi} は、ポーズ直前の発話単位における末尾から i 番目の音素の中心 20ms のフレームの正規化平均パワーである。また、p_{fi} は、ポーズ直後の発話単位における先頭から i 番目の音素の中心 20ms のフレームの正規化平均パワーである。正規化は、3.2.3 節と同じ要領で行った。P_b, P_f はそれぞれポーズの直前、直後のパワーを表わし、P_{bf} はその差を表わしている。

4.3.3 ポーズ長

ポーズ長を入力パラメータの一つとする。

4.4 音声データ

使用した講演音声データは、約 10 分のＮＨＫ論説番組「あすを読む」の 3 回分である。表 3 に用いたデータの内容、発話単位を示す。このうち、データ 1 とデータ 3 は 3.3 節で示したデータにも含まれている。講演音声の書き起こしは、人手で行なった。

4.5 複数パラメータの組み合わせ

入力パラメータとして、4.2 節、4.3 節で述べたように、言語情報 2, 基本周波数情報 3, パワー情報 3, ポーズ長 1 の合計 9 個がある。この 9 個のパラメータからいくつかを選択して、分類 2 進木の生成に実際に用いる入力パラメータセットとする。用いた 4 種類のパラメータセットを表 4 に示す。

4.6 評価結果

ポーズで区切られた発話単位の境界に対して、文境界であるかどうかを 2 進分類木で自動決定した。F 値で評価した結果を図 7 に示す。ここでオープンな評価では、3 つの講演音声データのうち 2 つのデータで 2 進分類木を学習し残りのデータで評価する処理を評価データを変えて 3 回行った結果を平均した。

表 4: パラメータの組合せ

パラメータセット	set1	set2	set3	set4
品詞 (前)	○	○	○	
品詞 (後)	○	○	○	
ポーズ		○	○	
基本周波数 (前)			○	○
基本周波数 (後)			○	○
基本周波数 (前-後)		○	○	○
パワー (前)			○	○
パワー (後)			○	○
パワー (前-後)		○	○	○

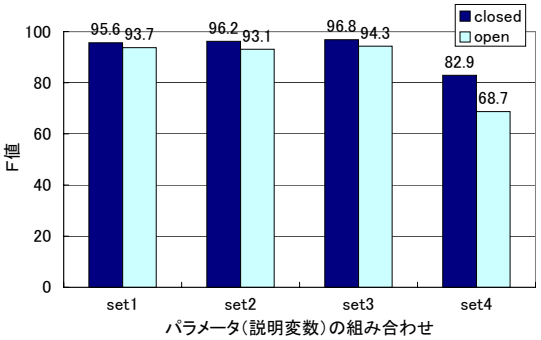
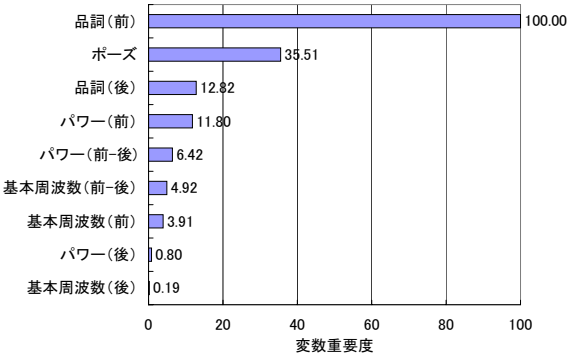


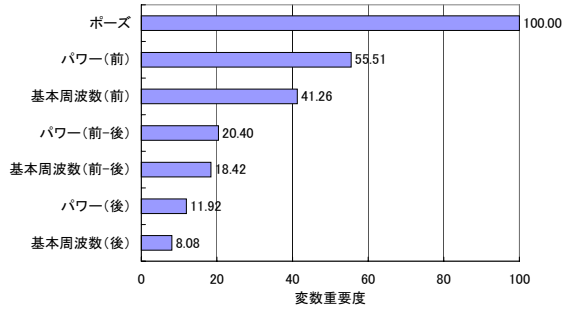
図 7: 文境界の自動決定の F 値による評価

図 7 から、全てのパラメータを用いる set4 において最も高い F 値 94.4% を得た。しかし、韻律情報を用いない set1 においても 93.7% の F 値が得られており、今回用いた講演音声では、言語情報だけを用いることでかなりの高精度で文境界の決定が行われており、韻律情報を利用することによる精度の改善はわずかであった。論説番組の音声では、文末が「～です」「～ます」で表現されることが多く、品詞情報によって文末がほぼ決定可能であったためである。言語情報を用いない set4 では、F 値が大きく減少し 68.7% となったが、韻律情報だけを用いることによっても、ある程度の文境界決定が行えることを示している。

2 進分類木を生成した時の変数重要度 [17] を set3 および set4 のパラメータセットに対して図 8 に示す。全てのパラメータを用いた (a) の set3 では、品詞の重要度が大きい。また、韻律情報に関しては、ポーズの重要度が大きく、先行発話単位末のパワーも重要であることがわかった。



(a) set3



(b) set4

図 8: 2 進分類木における変数重要度

5 まとめ

重要文抽出における講演音声の自動要約を実現するために、文境界さらに文重要度の自動決定において韻律情報を利用する手法について検討した。文重要度の決定では、連続音声認識システムを用いて講演内容の書き起こしテキストを作成した場合に、韻律情報を利用する効果が大きくなることが示された。文境界の決定では、今回用いた講演音声データでは、言語情報だけで精度の高い文境界の決定が行えたこともあり、韻律情報を利用する十分な効果は得られなかったものの、文境界の決定に有効な韻律情報として、ポーズ長、パワーの重要性が示された。今後の課題として、韻律情報の利用方法の高度化、要約精度の評価手法の再検討や他の講演音声データでの評価などが挙げられる。

参考文献

[1] I. Mani and M. Maybury : "Adavances in Automatic Text Summarization", The MIT Press (1999).

[2] 奥村学, 望月源 : "テキストを自動的に要約する技術 —第 1 回— テキスト中の重要な文を抜き出す", コンピュータサイエンス誌 bit2 月号, 共立出版, pp.37-42 (2000).

[3] 中川聖一 : "音声認識研究の動向", 電子情報通信学会論文誌, **J83-DII**, 2, pp.433–457 (2003).

[4] 笠原力弥, 山下洋一 : "講演音声における重要文と韻律的特徴の関係", 情報処理学会研究報告, SLP-35-5 (2001).

[5] 井上章, 三上貴由, 山下洋一 : "複数の韻律パラメータを用いた音声要約のための文重要度予測", 日本音響学会春季研究発表会講演論文集, 2-4-6, pp.69–70 (2003).

[6] 堀智織, 古井貞熙 : "単語抽出による音声要約文生成法とその評価", 電子情報通信学会論文誌, **J85-DII**, 2, pp.200–209 (2002).

[7] 小林聡, 吉川裕規, 中川 聖一 : "表層情報と韻律情報を利用した講演音声の要約", 情報処理学会研究報告, SLP-43-7 (2002).

[8] 北出祐, 南條浩輝, 河原達也, 奥乃博 : "談話標識と話題語に基づく統計的尺度による講演からの重要文抽出", 情報処理学会研究報告, SLP-46-2 (2003).

[9] S. R. Maskey and J. Hirschberg : "Automatic Summarization of Broadcast News using Structural Features", Proc. of Eurospeech 2003, pp.1173-1176 (2003).

[10] B. Wrede and E. Shriberg : "Spotting "Hot Spots" in Meetings: Human Judgments and Prosodic Cues", Proc. of Eurospeech 2003, pp.2805-2808 (2003).

[11] http://www.tufs.ac.jp/ts/personal/motizuki/software/posumcl/

[12] http://www.entropic.com/

[13] http://htk.eng.cam.ac.uk/

[14] http://www.lang.astem.or.jp/CSRC/

[15] http://julius.sourceforge.jp/

[16] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone : "Classification and Regression Trees", Wadsworth & Brooks (1984).

[17] 大滝厚, 堀江宥治, D. Steinberg : "応用 2 進木解析法 —CART による—", 日科技連 (1998).

[18] http://chasen.naist.jp/hiki/ChaSen/