

表層的言語情報と韻律情報を用いた 講演音声の重要文抽出

小林 聡^{††} 山口 優[†] 中川 聖一[†]

音声情報は記録が容易であるが、記録後の参照は必ずしも容易ではない。本論文では、音声の自動抽出要約を目指し、表層的言語情報による重要文抽出結果と人間による結果との比較を行ない、音声の抽出要約に有用と思われる表層情報を得た。また、韻律情報として F_0 とパワーに着目し、人間による結果との比較を行った。最後に表層的言語情報と韻律情報を組み合わせた自動抽出要約システムを作成し、そのシステムによって得られる抽出要約と人間による抽出要約との比較、聴取実験を行った。結果として F 値=0.599, κ 値=0.420 と良好な結果を得た。

キーワード: 講演音声, 要約, 表層的言語情報, 韻律情報, 手がかり語, ポーズ

Extractive Summarization of Lecture-style Spontaneous Speech Based on Linguistic Surface and Prosodic Information

SATOSHI KOBAYASHI^{††}, MASARU YAMAGUCHI[†] and SEIICHI NAKAGAWA[†]

Although it is easy to record speech, it is not easy to refer to audio recordings. If it is able to index or summarize audio recordings, referring to them would become easier. In this paper, we aim at extracting automatically the summarization of spoken lectures. For this purpose, at first we compared results of summarization extracted by human subjects. We found large differences with every subject. Then we investigated relations between linguistic surface information and human's results, and we obtained useful surface linguistic information. Next, we summarized spoken lectures based on this information, and compared them with human's results. Additionally, we focused on prosodic features; F_0 and power. We conducted the same experiments on them. Lastly, we combined linguistic surface information and prosodic information. As a result, we obtained a better F -measure(0.599) and κ -value(0.420), comparable with human's results.

KeyWords: *Lecture-style Speech, Summarization, Linguistic Surface Information, Prosodic Information, Cue Word, Pause*

1 はじめに

音声情報は、その記録は容易であるが、記録後の参照は必ずしも容易ではなく、そのためにはインデクス化や文書化しておく必要がある (Zechner 2002; 長谷川, 秋田, 河原 2001; 下岡, 河原, 奥乃 2002). 音声情報をインデクス化したりそのまま要約することが可能となれば、音声

[†] 豊橋技術科学大学情報工学系, Information and Computer Sciences, Toyohashi University of Technology

^{††} 島根大学総合情報処理センター, General Information Processing Center, Shimane University

情報への参照も容易かつ簡便なものとなる。従来、テキストからの自動的な重要文抽出や要約の研究は行なわれている(奥村, 難波 1999; 野畑, 関根, 井佐原, R. 2002a; 野畑, 関根, 山元, 井佐原 2002b)。テキスト要約の主な素性には‘手がかり語’, ‘タイトル語’, ‘キーワード’, ‘文の位置’がある(Edmundson 1969)。手がかり語はコーパスの訓練部分集合から抽出された単語であり, タイトル語はタイトル, 副題, 見出し中の単語で, 文書の文中にある単語である。キーワードは手がかり語と異なり文書固有のもので, 単語頻度を求め, その単語頻度を重みにもつキーワードとして抽出される。文の位置はテキスト中の順序位置をもとに文に重みを割り当てる。特に, 最初または最後の段落に出現する文, あるいは段落の最初または最後の文の場合に正の重みを割り当ててものとなっている(Inderjeet Mani, 奥村, 難波, 植田 2003)。例えば, 新聞の社説の要約に関しては, 野畑らは主に文の位置や文長, *tf-idf* を用いて, 要約率 30%において *F* 値が 0.256 と報告している(野畑, 関根, 井佐原 2003)。また, テキスト自動要約タスク(TSC) データ(TSC 2001)における新聞記事の重要文抽出では, 要約率 30%で *F* 値が 0.435 という値を得ている。

しかし, 音声を対象とした要約研究は少ない。講演音声の効率的利用に関しては, 長谷川らによる研究(長谷川他 2001)がある。これは, 談話標識となる単語を自動的に抽出するものであるが, 文境界の検出などに多量の学習データを必要とし, またインデキシングされるのみであるため講演内容の概略の把握が難しい。文境界を人手もしくは自動的に検出し重要文抽出を行う研究もなされており(野畑他 2003), 日本語話し言葉コーパス(CSJ)の講演音声(前川 2001)において正しい文境界を予め与えた場合の重要文抽出結果は *F* 値で 0.368 と報告している。しかし, 被験者間の重要文の判断はばらつきが大きく, それらを統合して正解データを作成することが困難なため, 各被験者の重要文抽出結果を個々に正解として扱っている。また北出らも CSJ の講演音声を用い, 談話標識と話題語に基づいた重要文抽出を行っており, 再現率 0.519, 適合率 0.369, *F* 値 0.431 という結果を得ている(北出, 南条, 河原, 奥乃 2003)。新中らは音声認識を利用した CSJ の講演音声について, 単語 *trigram* による言語スコア, 単語の出現頻度に基づく単語重要度スコア, 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ペナルティを与える信頼度スコアを用いて重要個所の抜き出し法を検討し, 人間の要約結果(単語列)を正解とみなし, 単語レベルでの比較で約 45%の要約(単語)正解精度を得ている(新中, 菊地, 岩野, 古井 2003b)。なお, 新中らは単語単位, 文単位, フィラー単位での抽出を試みており, 要約された音声の聴取実験では文単位による要約が聞き易さで優れており, テキスト要約で行なわれている一文内の要約(単語単位)の抽出要約では聞き易さ, 要約としてのふさわしさの点で文単位のものよりも悪い結果となっている。なお, 我々は音声要約としては, 以前から文単位で行っている(小林, 吉川, 中川 2002; 吉川, 小林, 中川 2003)。

また, ニュース音声を対象とした要約では, Waibel らによる MMR(Maximum Marginal

Relevance)(Carbonell, Geng, and Goldstein 1997) 手法を用いた要約システム (Zechner 2003; Waibel, Bett, and Finke 1998) がある. MMR 法とは冗長性を削減する手法として情報検索の分野で用いられているもので, 検索質問に対するパッセージの関連度とパッセージとの間の類似度から同義部分と相違部分を検出でき, 冗長性を削減することができる. MMR 法を用いることで Waibel らの従来の要約システムより良い要約が可能なが分かっている. また, Zechner らはニュース音声から電話での会話や会議などの討論に要約のタスクを拡張している. この要約システムは MMR 手法をベースにしており, 要約文の文法性の維持と自動的に文の境界を挿入するための機械学習を行っている. その結果, このシステムは会話の要約において有効なことが示されている (Zechner 2003).

また, 笠原や三上らにより, NHK の「あすを読む」を対象とした韻律的特徴と重要文との関係が検討されている (笠原, 山下 2001; 三上, 井上, 山下 2002). 「あすを読む」はニュースの解説番組であり, 話す内容や話す順序が推敲されているが, ニュース音声のような原稿の読み上げ音声というよりも, 原稿に沿った話し言葉に近い音声である. 「あすを読む」に関して, 文単位ではパワーに関して文の重要度とやや強い相関が報告されているが, 韻律情報だけからの要約は難しい. そこで, 井上らは, ピッチ情報と言語情報を用い, 言語情報の利用により要約精度が向上することを示している (井上, 三上, 山下 2002). 「あすを読む」に関しては伊藤らによっても研究されており, 要約率約 30% において, 3 人の被験者間の κ 値が 0.42 となっている. また, 言語情報や文の位置などを利用して重要文抽出を行っており, 文の位置では適合率が 0.456, *tf · idf* で適合率が 0.395 などとなっている (伊藤, 松本, 谷田, 柏岡, 田中 2001).

我々も, 「あすを読む」が主題, 主張が明らかであり重要文抽出が容易であると考え音声要約を試みたが, 被験者による重要文の選び方にばらつきが大きく, 被験者間の一致率が 0.506, κ 値が 0.222 という低い値となった (山口, 中川 2005). 被験者が内省したところ, 同じ内容を 2 度説明する箇所, すなわち, はじめに概要を説明しその後詳細を述べる箇所があり, 概要と詳細のどちらを選ぶか迷ったという意見があった. そのような箇所における各被験者の判断の違いが被験者間のばらつきを大きくしているものと思われる.

本論文では, 「話し言葉工学」プロジェクト (CSJ) コーパスの講演音声を, より多くの表層的言語情報と韻律情報を用いて自動的に重要文を抽出して要約する方法を検討し, 人手による要約と比較した結果について述べる. 本研究では, 約 12 分程度の講演音声を約 4 分程度に要約すること (要約率 33%) を目指し, 重要文抽出に基づく要約部分を音声で提示することを目的とする. 講演音声の認識結果は重要文抽出に用いるが, ユーザに提示しないので認識誤りはさほど問題は生じない. 一文中の言い換えや冗長部分の削除による要約は難しく要約率も小さい (要約率は 80% 前後) (三上, 増山, 中川 1999). また, 音声は削除した単語や文節の接続部分が不自然になることから (新中, 菊池, 古井 2003a) 一文中の要約は行っていない. 韻律情報では発話時間長, 表層的言語情報においては手がかり語や頻出単語などの特徴が重要文抽出において

有用であることを示す。また、表層的言語情報と韻律情報の複数の特徴を組み合わせることで人手による書き起こしデータに対して F 値=0.599, κ 値=0.420, 音声認識による自動書き起こしデータに対して F 値=0.574, κ 値=0.384 という良好な結果が得られたことを述べる。最後に、自動抽出された文に対応する音声をつなぎ合わせ、作成した要約音声の聴取実験について述べる。

以下 2 節では、実験に利用した試料や実験条件、本論文で用いる評価尺度について述べる。3 節では、人手の要約における被験者間の重要文の一致の度合について述べる。4 節では、韻律情報を利用した重要文抽出の手法について述べる。5 節では、表層的言語情報を利用した重要文抽出の手法について述べる。6 節では、抽出された要約文候補の棄却可能な特徴について説明する。7 節では、表層的言語情報と韻律情報を組み合わせ実験について説明する。8 節では、評価実験結果と作成した要約の聴取実験について述べる。9 節では、まとめとともに今後の展望について述べる。

2 音声試料と評価尺度

2.1 音声試料と書き起こし

本研究で利用した音声試料について説明する。本研究では「話し言葉工学」プロジェクトにより提供されている日本語話し言葉コーパスの内、特に日本音響学会における 5 つの講演音声 (男性話者 5 名) を対象として要約実験を行なった。表 1 に講演音声の諸元を記す。講演音声の書き起こしの例を図 1 に挙げる。書き起こしには「えー」などのフィラー情報が含まれている。

なお、以下で用いる「文」の認定についてはコーパスの書き起こしテキストに従った (原則として 200msec. 以上のポーズにより分割)。

表 1 講演音声の諸元

講演音声	文の総数	時間長
a01m0037	184	13 分 35 秒
a01m0086	313	14 分 04 秒
a01m0088	281	14 分 02 秒
a01m0124	185	13 分 14 秒
a01m0157	333	11 分 22 秒
平均	259.2	13 分 15 秒

また、音声認識結果を用いて自動書き起こししたデータを用いても実験を行なっている。使用した大語彙日本語連続音声認識システムは、IPA「日本語ディクテーション基本ソフトウェアの開発」プロジェクトから提供されている Julius (河原, 李, 小林, 武田, 峯松, 嵯峨山, 伊

(F えー) ですがども (F えー)
 欠点としまして
 (F えー) 通常 (F えー)
 ある回帰クラスに対し共通の適応する訳ですがども
 その回帰クラスの実質的な
 (F あの) 含まれる分布数が少なくなりますと
 (F えー) その回帰行列の次元より以下になりますと不良条件になりまして
 (F えー) 推定ができないという (F ま) そういう問題があります

図 1 書き起こしの例

藤, 伊藤, 山本, 山田, 宇津呂, 鹿野 2000) および, 豊橋技術科学大学 工学部 情報工学系 中川研究室で開発された SPOJUS(赤松, 花井, 甲斐, 峯松, 中川 1998) を利用した. この2つの音声認識システムの出力結果の共通部分を音声認識結果として, 重要文抽出に用いている. なお, Julius と SPOJUS の音声認識システムの認識結果の共通部分をとることにより高い信頼度を得ることができる (宇津呂, 西崎, 小玉, 中川 2003; 宇津呂, 小玉, 渡邊, 西崎, 中川 2004).

2.2 評価尺度

本研究では評価尺度として κ 値, 適合率, F 値を用いた. 以下に各評価尺度について説明する.

- κ 値

κ 値 (Siegel and Castellan 1988) とは2者の判定の一致度を, 偶然の一致を考慮して調整した指標であり, 次式により定義されている.

$$\kappa = \frac{P(A) - P(E)}{1 - P(E)} \quad (1)$$

$$P(A) = \frac{\text{重要文の一致数} + \text{非重要文の一致数}}{\text{文の総数}} \quad (2)$$

$$P(E) = \text{重要文の偶然の一致率} + \text{非重要文の偶然の一致率} \quad (3)$$

$$\text{重要文の偶然の一致率} = \frac{A \text{ が重要と判定した文}}{\text{文の総数}} \times \frac{B \text{ が重要と判定した文}}{\text{文の総数}} \quad (4)$$

$$\text{非重要文の偶然の一致率} = \frac{A \text{ が非重要と判定した文}}{\text{文の総数}} \times \frac{B \text{ が非重要と判定した文}}{\text{文の総数}} \quad (5)$$

被験者 A が重要と判定した文と被験者 B が重要と判定した文，そして重要文の一致数が決まれば，非重要文の一致数は一意に決まる．また，A が重要と判定した文数が B が重要と判断した文数と近くなるほど κ 値は高くなるという特性があり， κ 値の一致度の解釈は表 2 のようになる (Carletta, Isard, Kowtko, Doherty-Sneddon, and Anderson 1997)．なお κ 値については $0.4 < \kappa < 0.6$ を fair, $0.6 < \kappa < 0.75$ を good, $0.75 < \kappa$ を excellent とする指標も提案されている (Fleiss 1971)．

表 2 κ 値の一致度の解釈

κ 値	一致度
< 0	Poor
0 - 0.2	Slight
0.21 - 0.4	Fair
0.41 - 0.6	Moderate
0.61 - 0.8	Substantial
0.81 - 1	Near Perfect

- 一致率

一致率とは，人間が抽出した文集合 (正解文集合と見なす) を H とし，機械が抽出した文集合を M とすると，以下の二式が考えられる．(6) 式は機械から見た人間との適合率 (Precision)，(7) 式は人間から見た機械との再現率 (Recall) を表している (野畑他 2003)．本研究では (6) 式の適合率を用いる．

$$\text{適合率 (Precision)} = \frac{|M \cap H|}{|M|} \quad (6)$$

$$\text{再現率 (Recall)} = \frac{|M \cap H|}{|H|} \quad (7)$$

- F 値

F 値は適合率と再現率の調和平均を示し，(8) 式により与えられる．

$$F \text{ 値} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (8)$$

3 人間による重要文抽出

本研究で用いた講演音声は学会発表時のもので、講演内容が極めて専門性が高いため、人間による重要文抽出における被験者は、対象とする講演の内容を理解できる音声研究の専門家6名(A-F)とした。各被験者は対象とする講演について、書き起こしを読み重要文の抽出要約を行った(吉川他 2003)。実際の聴講時のように講演音声を聴きながら逐次的に重要文を抽出する被験者実験も行なっている(小林他 2002)が、書き起こしテキストの場合とほぼ同等の結果である。

本研究では要約の長さを講演の3分の1(要約率33%)を目標としており、各被験者に抽出文数は講演全体の3分の1程度の分量になるように指示した。そのため、被験者により抽出文数は異なる。また、ここで被験者5人(A-E)中3人以上が重要と判断した文を人間の要約とし **man3** と呼ぶことにする。**man3** をとることで、被験者のばらつきが小さくなり、より安定に重要な文を得ることが可能となる。

はじめに被験者間の重要文の一致の度合を見るため、被験者間の κ 値と適合率および各被験者と人間(その被験者を除いた5名による **man3**)の間の κ 値と適合率を求め、それぞれ表3(a)と表3(b)に示す。また、表3(b)の“A-man3”は“AとAを除いた5名による **man3**”を意味する。表3(a)より被験者間の適合率は0.550、 κ 値は0.286とそれほど高い値とは言えない。なお、我々の行ったNHKの「あすを読む」における同様の5人による被験者実験の結果、被験者間の適合率が0.506、 κ 値が0.222と低く、「あすを読む」は講演音声(CSJ)と比べ被験者間の重要文の判定に大きなばらつきが見られた(山口, 中川 2005)。表3(b)より、各被験者と **man3** の κ 値は0.407と比較的高い値となった。この値が自動要約手法の目標値となる。

表3 被験者間の一致の度合

(a) 被験者間の一致の度合

	A-B	A-C	A-D	A-E	B-C	B-D	B-E	C-D	C-E	D-E	平均
適合率	0.302	0.429	0.566	0.491	0.618	0.631	0.717	0.616	0.581	0.545	0.550
κ 値	0.200	0.089	0.245	0.331	0.288	0.246	0.370	0.339	0.358	0.393	0.286

(b) 被験者とその被験者を除いた **man3** の間の一致の度合

	A-man3	B-man3	C-man3	D-man3	E-man3	F-man3	平均
適合率	0.793	0.396	0.583	0.679	0.639	0.680	0.629
κ 値	0.339	0.348	0.392	0.413	0.464	0.484	0.407

4 韻律情報を利用した重要文抽出

音声には韻律的な情報を含んでおり、この韻律情報を用いて抽出要約の精度の向上を図ることができると考えられる。韻律情報として F_0 (基本周波数)、パワー、発話時間長、ポーズを用

いた。これらについて様々な特徴パラメータを抽出したが、以下で述べる特徴が比較的有用であった(吉川他 2003)。

4.1 $F0$

各文の $F0$ の平均と標準偏差を求め、“平均 + 標準偏差”以上の文を抽出する ($F0_{avg}$)。 $F0$ の最初の 30 フレーム (1 フレーム 10msec.) の傾きの平均と標準偏差を求め、“平均 + 標準偏差”以上の値を傾きが正と定め、傾きが正の文を抽出 ($F0_{grd+}$)、同様に“平均 - 標準偏差”以下の値を傾きが負と定め、傾きが負の文を抽出 ($F0_{grd-}$) する。

4.2 パワー

$F0$ と同様の手法により POW_{avg} と POW_{grd+} , POW_{grd-} を求め抽出を行う。

4.3 発話時間長

各文の発話時間長の平均を求め、平均の高い文から上位 3 分の 1 抽出する (LEN)。

4.4 ポーズ

1 秒以上のポーズを長いポーズと定義する。その長いポーズを境に話題の転換が考えられ、この話題転換の境界付近に重要文の存在がいくつか確認された。長いポーズの前 5 文 ($PAUSE_{pre5}$) と、長いポーズの後 5 文 ($PAUSE_{pos5}$) を抽出する。また、講演全体から長いポーズのベスト 10 を求め、長いポーズベスト 10 の前 5 文 ($PAUSE_B10_{pre5}$) と長いポーズベスト 10 の後 5 文 ($PAUSE_B10_{pos5}$) を抽出する。

5 表層的言語情報を利用した重要文抽出

表層的言語情報とは手がかり語や文の位置といった情報である。本研究では手がかり語、頻出単語、 tf ベース、文の位置の各特徴を書き起こしテキストもしくは音声認識結果を基に利用した(吉川他 2003; 伊藤他 2001; 野畑他 2003)。それぞれの詳細について以下で説明する。

5.1 手がかり語

手がかり語とはその前後に重要文があるであろうという観点から定義されている。手がかり語としては“与えております”や“いう結果になってます”、“結果がこれです”など全 75 個を定

めた。手がかり語は音声認識 (甲斐, 廣瀬, 中川 1999) によって得られた結果を用い, 手がかり語の認識率を向上させるため言語モデルの中の手がかり語に関する 1-gram, 2-gram, 3-gram の値を 3 倍にしている。これにより, 全体の認識精度は下がるものの (単語認識率 52% から 49%), 手がかり語の認識率は 76% から 90% にまで向上できた。なお, 手がかり語の沸き出し誤りは生じなかった。

それぞれの手がかり語ごとに手がかり語を含む文とその前 5 文, 手がかり語を含む文とその後 5 文, および手がかり語を含む文とその前後 5 文を抽出した文と人間による要約結果との適合率を求め, その値が 40% 以上になる手がかり語を表 4 に示すように分類した。手がかり語を含む文とその前 5 文 (表 4 (a)) (CUE_{pre5}), 手がかり語を含む文とその後 5 文 (表 4 (b)) (CUE_{pos5}), 手がかり語を含む文とその前後 5 文 (表 4 (c)) (CUE_{in5}) を抽出する。これらによって抽出された文の集合を CUE_{all} とする。

表 4 手がかり語の細分類

(a) 手がかり語の前 5 文を抽出する手がかり語の例 (全 34 個)

いう結果になっています 結果を発表します	重要になってくる 示しました	そういう問題があります 検証できました
-------------------------	-------------------	------------------------

(b) 手がかり語の後 5 文を抽出する手がかり語の例 (全 17 個)

いうことが課題 報告されております	研究の背景 評価を行い	本研究のねらい 本日の発表内容
----------------------	----------------	--------------------

(c) 手がかり語の前後 5 文を抽出する手がかり語の例 (全 13 個)

最後にまとめます 削減ができております	提案いたしました 認められるが	考えております 与えております
------------------------	--------------------	--------------------

5.2 頻出単語

頻出単語とは書き起こしテキストもしくは音声認識結果中に何度も現れる単語を意味する。頻出単語の検出には形態素解析システム『茶釜』(松本, 北内, 山下, 平野, 松田, 高岡, 浅原 2003) を使用し, 不要語やフィラー, 数詞を除いた一般名詞のみを頻出単語として定義する。類似したものとして情報検索で有用な $tf \cdot idf$ 法 (Salton and Yang 1973) があり, tf では語頻度 (語の出現回数) を文のスコアとして文抽出を行う。しかし, 長い文にはより多くの単語が含まれるため, 長い文が抽出されやすいという傾向が見られた。そこで, 頻出単語においては, 1 文中にその単語が 2 語以上含まれている文を重要文とみなし文抽出を行う。書き起こしもしくは音声認識結果において頻出単語の出現回数を調べ, 頻出単語を含む文を全体の 3 分の 1 程度と

なるよう単語数の閾値を決め文を抽出する ($WORD_{rpt}$). なお, 書き起こしを用いた場合の単語数の閾値は5つの講演において7, 8, 9, 3, 22単語であり, 表5に頻出単語の例を示す.

表5 頻出単語の例

a01m0037		a01m0086		a01m0088		a01m0124		a01m0157	
頻出単語	出現回数	頻出単語	出現回数	頻出単語	出現回数	頻出単語	出現回数	頻出単語	出現回数
韻律	23	振動	55	回帰	26	単語	33	カテゴリー	22
合成	17	口蓋	31	学習	19	クラス	28	言語	17
音声	14	適応	17	データー	18	グラム	28	非常	12
ベクトル	13	ランダム	14	行列	18			状態	11
ピッチ	11	動作	14	分布	14			読点	11
モーラ	11	発声	13	クラス	13			意味	10
自然	9	補償	13	ランク	13			韻律	10
		条件	12	閾	12			情報	10
				適応	10			予測	10
								モデル	9
								解析	8
								パラメーター	6
								制御	6
								生成	6
								用法	6

5.3 tf

tf (term frequency) とは文書中に現れる単語の出現頻度のことである. 今回 tf ベースでの重要文抽出には自動要約器 Posum(望月 2002) を用いた. Posum では, 書き起こしや音声認識結果のテキストを茶筌により形態素解析した結果から, 名詞, 動詞, 形容詞と未定義語の各単語を情報として用い, 各文の重要度を計算している. 各文の重要度は, その文に出現する単語の重みを, その出現頻度 (tf) や出現文書頻度の逆数 (inverse document frequency : idf) を用いた tf や $tf \cdot idf$ によって計算し, 足し合わせることで計算する (長尾, 宇津呂, 嶋津, 匂坂, 井口, 片寄 2000). 本実験では不要語やフィラーを除去しカウントする品詞を名詞のみとして tf ベースで全体の3分の1程度となるよう抽出を行う (TF). 各単語の重み w_i は (9) 式で与えられる. ここで tf_i は単語 i の文書内の出現頻度を示す. $tf \cdot idf$ についても実験を行ったが tf の場合と適合率がさほど変わらず, 今回の実験では簡便性から tf で行った. なお, 単語数に対する正規化を行っても行わなくても同じ結果になるので正規化は行っていない. TF による手法を本研究のベースラインとする.

$$w_i = tf_i \quad (9)$$

5.4 文の位置

文の位置情報として, 各講演のはじめの10文 ($LEAD_{10}$), 最後の10文 ($TAIL_{10}$) の抽出を行う.

6 重要文候補の棄却可能な特徴

これまで韻律情報と表層的言語情報を用いた重要文抽出の手法について述べてきたが, 非重要文を棄却することでより良い要約が可能となると考えられる. 様々な特徴を検討した結果, 以下に棄却可能な特徴を示す.

6.1 $F0$ / パワーの平均の低い文の削除

各文の $F0$ とパワーの平均の低い文は重要文である可能性が低いと考えられる. 各文の $F0$ / パワーの平均値を 1 講演全ての文に対して求め, その“平均-標準偏差”以下を $F0$ の低い文 ($\#F0_{low}$), パワーの低い文 ($\#POW_{low}$) と定義する. なお, 各文の前後の無音区間は無視して考える.

6.2 発話時間の短い文の削除

$F0$ / パワーの低い文と同様に, 発話時間の短い文は重要文である可能性は低いと考えられる. 各文の発話時間の平均値を 1 講演全ての文に対して求め, その“平均-標準偏差”以下を発話時間の短い文 ($\#LEN_{short}$) と定義し抽出する.

6.3 長いポーズに挟まれた数文の削除

前節までに, 長いポーズの前/後 5 文の抽出について述べた. その長いポーズの前後 5 文を見ると, その中に別の長いポーズの存在が確認された. その長いポーズ間の数文はデモの最中であったり, 言い誤って言葉に詰まっているところが見受けられ (図 2), 重要文である可能性は低いと考えられる. ここでは長いポーズベスト 10 の間に挟まれた 5 文以内の文 ($BETWEEN_{lp-lp}$) を棄却可能対象とし抽出する.

7 韻律情報と表層的言語情報を組み合わせた要約

書き起こしと音声認識による自動書き起こしに対して, 韻律情報と表層的言語情報の中で比較的結果の良かった特徴を組み合わせて要約を行なう. 組み合わせに用いる特徴は前節までに説明した韻律情報と表層的言語情報の中で, 適合率が 40% 以上の値を得ることのできたものとした. 特徴 F_k により重要と推定された i 番目の文 S_i に対するスコア $Score_{F_k}()$ を (10) 式のように定義する.

$$Score_{F_k}(S_i) (1 \leq i \leq n) = \begin{cases} 1 & (\text{重要文として抽出された文}) \\ 0 & (\text{それ以外の文}) \end{cases} \quad (10)$$

それに基づきまして (F えー) ノルムを最小にするという拘束つきで
 (F えー) 解を求める
 そういう (F えー) 検討をした結果を (F えー) 発表いたします
 <Long Pause (A)>
 まず最初に (F えー)
 不良条件が発生するという (? すんませ)(? あん) そのこのところの
 説明をいたしますと
 <Long Pause (B)>
 (F ま) 通常 (F えー) 学習データーに対しまして
 (F えー) 尤度最大ということで (F えー)

図 2 抽出する位置の例 (太字：削除対象文)

ここで n は文書中の総文数を示す.

また, 棄却可能な特徴 D_p により非重要と推定された文 S_i に対するスコア $Score_{D_p}()$ を (11) 式のように定義する.

$$Score_{D_p}(S_i) (1 \leq i \leq n) = \begin{cases} 1 & \text{(非重要文として抽出された文)} \\ 0 & \text{(それ以外の文)} \end{cases} \quad (11)$$

各特徴の値 ($Score_{F_k}()$, $Score_{D_p}()$) に各特徴の寄与度 (α_k, β_p) を掛け合わせたものの線形和をとって各文 S_i のスコアとする.

$$TotalScore(S_i) = \sum_k \alpha_k Score_{F_k}(S_i) + \sum_p \beta_p Score_{D_p}(S_i) \quad (12)$$

ここで, 寄与度 α_k は 0~0.6 まで 0.2 刻み, β_p は 0 か $-\infty$ とした. $TotalScore()$ の高い文から順に全体の 3 分の 1 となるように抽出を行い, 同スコアの場合には講演の後ろを重視し, 後ろの文から順に抽出する. 寄与度の最適値は 4 講演において共通な値を求める.

また, 菊池ら (菊池, 古井, 堀 2002) も表層情報に基づき, 複数の特徴の重み付き線形和で文の重要度を定義し, 要約率が 50% と 70% の条件で重要文抽出を行っている. 要約率が 70% のときは複数の特徴を組み合わせることで要約精度の改善が見られる. しかし, 要約率が 50% においては複数の特徴を組み合わせることによる要約精度の改善はあまり得られていない. また,

どちらの要約率の場合においても, 言語スコア, 文の重要度スコア及び信頼度スコアを組み合わせた重要文抽出手法は言語スコアを用いない場合と比べて要約精度の改善があまり見られない. 菊池らは韻律情報を用いていないが, 我々は表層的言語情報と韻律情報を組み合わせ, 重要文抽出における韻律情報の有効性を示す.

8 実験結果

8.1 韻律情報と表層的言語情報における重要文抽出結果

韻律情報を利用した重要文抽出の結果を表 6 に示す. 表 6 は韻律情報を用いて抽出したシステムの結果と人間 (man3) の結果との間の適合率と κ 値と F 値である. また, 表 7 に表層的言語情報を用いた際の結果, 表 8 に棄却可能な特徴による結果を示す. 表 6 から韻律情報では $F0_{grd-}$ と POW_{grd-} , LEN が比較的高い適合率を得ていることが分かる. しかし, 偶然の適合率は約 0.33 であり, LEN を除いては決して高い適合率とは言い難い.

表層的言語においては (表 7), 手がかり語 ($CUE_{pre5}, CUE_{pos5}, CUE_{in5}$) と頻出単語 ($WORD_{rpt}$) が比較的高い値を得ていることが分かる. なお, $WORD_{rpt}$ と TF については音声認識結果 (Julius と SPOJUS の出力結果の共通部分) を用いた際の結果であり, 手がかり語については前述の通りである. なお, CUE_{in5} の a01m0157 については手がかり語が存在しなかった.

また, 棄却可能な特徴については (表 8) LEN_{short} と $BETWEEN_{lp-lp}$ の適合率が低く, これらの文は重要文である可能性は低く, 削除することにより適合率と κ 値の上昇が望める.

8.2 組み合わせ実験結果

4 つの講演音声 (a01m0037, a01m0086, a01m0088, a01m0124) により求めた寄与度の最適値を表 9 に示す¹. ‘#’ は削除する特徴であり, 表の空欄は寄与度が 0 を意味している. 表 9 から, 講演の最後の 10 文 (結論部分) や長いポーズの前 10 文, 頻出単語などが重要文抽出に寄与していることがわかる. 表 10 に κ 値の最も良かった組み合わせの一致の度合いを示した. 5 講演の平均で κ 値=0.420, 適合率=0.574, F 値=0.599 となり, ベースラインである TF と比較して高い値を得られた. これは, 表 3 (b) で示した man3 と被験者間の κ 値および適合率とほぼ同等であり, 人間並みの重要文抽出ができたと言える. また, 音声認識による自動書き起こしデータを用いた場合でも, 人手による書き起こしデータとほぼ同等の結果が得られた.

また, 組み合わせ実験の最良の結果である適合率 0.574 は手がかり語 (CUE_{in5}) を用いた場合の適合率 0.543 と近い値であるが, 一般に抽出率が小さいほど適合率は大きくなる性質があるので直接比較できない. 手がかり語による文の抽出率は約 13 % であり, 抽出率約 13 % の場合

¹ 講演音声 a01m0157 に対しては, 寄与度に関してオープン実験である.

において組み合わせ実験を行った際の適合率は図 3 に示すように 0.624 であった。このことから両者間の適合率には有意な差があると考えられる（危険率 5% で統計的に有意）。また、他の単独手法である *tf* (ベースライン) や頻出単語による適合率と組み合わせの適合率の差も大きいことがわかる。比較的结果の良かった特徴と、組み合わせ実験（書き起こし）における κ 値と適合率、 F 値の 5 講演の平均を図 4 に示す。図 4 より、複数の特徴を組み合わせることで κ 値と適合率、 F 値ともに良い値が得られていることが分かる。

また、最も κ 値の良かった組み合わせパターンと寄与度 (表 9) について、韻律情報の寄与度が零 (韻律情報 ($PAUSE_B10_{pre5}$, $\#LEN_{short}$) を用いない) にしたときの結果を表 11 に示す。表 10 と比較して、人手での書き起こしデータと音声認識による自動書き起こしデータのどちらを用いた場合も κ 値、適合率、 F 値が大幅に低い値となっており、韻律情報が重要文抽出に有用であることが分かる。

表 6 韻律情報における重要文抽出結果

	κ 値	適合率	F 値	κ 値	適合率	F 値	κ 値	適合率	F 値
	$F0_{avg}$			$F0_{grd+}$			$F0_{grd-}$		
a01m0037	-0.155	0.160	0.080	-0.017	0.333	0.071	0.108	0.632	0.255
a01m0086	-0.071	0.170	0.129	-0.028	0.176	0.064	0.035	0.294	0.180
a01m0088	-0.063	0.188	0.109	-0.076	0.105	0.041	0.024	0.320	0.155
a01m0124	-0.085	0.174	0.104	0.021	0.333	0.145	0.089	0.444	0.222
a01m0157	-0.167	0.049	0.031	0.018	0.375	0.061	-0.003	0.267	0.178
平均	-0.108	0.148	0.091	-0.016	0.264	0.077	0.051	0.391	0.198
	POW_{avg}			POW_{grd+}			POW_{grd-}		
a01m0037	-0.106	0.133	0.044	-0.054	0.267	0.089	0.106	0.750	0.207
a01m0086	-0.077	0.159	0.116	0.112	0.385	0.259	-0.011	0.200	0.046
a01m0088	0.015	0.297	0.191	-0.013	0.258	0.147	0.037	0.444	0.092
a01m0124	-0.040	0.240	0.152	0.028	0.333	0.187	0.041	0.385	0.149
a01m0157	-0.152	0.087	0.059	0.012	0.300	0.109	0.003	0.276	0.134
平均	-0.072	0.183	0.112	0.017	0.309	0.158	0.035	0.411	0.126
	LEN			$PAUSE_{pre5}$			$PAUSE_{pos5}$		
a01m0037	0.212	0.557	0.500	0.063	0.452	0.409	0.109	0.473	0.470
a01m0086	0.130	0.327	0.376	0.065	0.305	0.265	-0.164	0.114	0.109
a01m0088	0.099	0.340	0.372	0.121	0.349	0.402	0.017	0.287	0.350
a01m0124	0.298	0.484	0.517	0.101	0.389	0.311	0.207	0.465	0.412
a01m0157	0.355	0.495	0.547	0.123	0.520	0.226	0.017	0.300	0.150
平均	0.219	0.441	0.462	0.095	0.404	0.323	0.037	0.328	0.298
	$PAUSE_B10_{pre5}$			$PAUSE_B10_{pos5}$					
a01m0037	0.094	0.500	0.359	0.070	0.476	0.342			
a01m0086	0.088	0.340	0.258	-0.090	0.149	0.113			
a01m0088	0.155	0.438	0.333	-0.107	0.167	0.127			
a01m0124	0.049	0.333	0.292	0.161	0.429	0.375			
a01m0157	0.185	0.489	0.336	0.059	0.340	0.233			
平均	0.114	0.410	0.316	0.019	0.312	0.238			

表 7 表層的言語情報における重要文抽出結果

	κ 値	適合率	F 値	κ 値	適合率	F 値	κ 値	適合率	F 値
	CUE_{pre5}			CUE_{pos5}			CUE_{in5}		
a01m0037	0.085	0.506	0.468	0.118	0.567	0.324	0.163	0.575	0.400
a01m0086	0.170	0.532	0.331	0.078	0.467	0.152	0.127	0.500	0.222
a01m0088	0.270	0.549	0.508	0.132	0.579	0.227	0.296	0.600	0.439
a01m0124	0.012	0.477	0.268	0.082	0.556	0.159	0.178	0.500	0.341
a01m0157	0.235	0.556	0.405	0.058	0.500	0.118	-	-	-
平均	0.154	0.522	0.396	0.094	0.531	0.196	0.191	0.543	0.351
	CUE_{all}			$LEAD_{10}$			$TAIL_{10}$		
a01m0037	0.135	0.467	0.549	-0.080	0.100	0.024	0.024	0.500	0.118
a01m0086	0.239	0.437	0.419	0.013	0.300	0.069	0.062	0.500	0.115
a01m0088	0.234	0.402	0.498	-0.043	0.100	0.023	0.030	0.400	0.091
a01m0124	0.289	0.476	0.513	0.003	0.300	0.094	0.072	0.500	0.156
a01m0157	0.242	0.479	0.422	-0.057	0.000	0.000	0.027	0.400	0.080
平均	0.228	0.452	0.480	-0.033	0.160	0.042	0.043	0.460	0.112
	$TF(\text{baseline})$			$WORD_{rpt}$					
a01m0037	0.328	0.639	0.573	0.387	0.677	0.613			
a01m0086	0.134	0.330	0.378	0.160	0.346	0.398			
a01m0088	0.154	0.376	0.409	0.232	0.426	0.465			
a01m0124	0.156	0.393	0.417	0.231	0.438	0.475			
a01m0157	0.445	0.555	0.610	0.427	0.546	0.596			
平均	0.243	0.459	0.478	0.287	0.487	0.509			

表 8 棄却可能な特徴における結果

	κ 値	適合率	F 値	κ 値	適合率	F 値	κ 値	適合率	F 値
	$F0_{low}$			POW_{low}			LEN_{short}		
a01m0037	-0.004	0.400	0.168	-0.206	0.080	0.040	-0.244	0.166	0.120
a01m0086	-0.002	0.243	0.158	-0.040	0.204	0.159	-0.178	0.061	0.048
a01m0088	-0.033	0.222	0.114	-0.025	0.242	0.144	-0.132	0.141	0.110
a01m0124	-0.066	0.214	0.146	-0.143	0.133	0.095	-0.181	0.097	0.071
a01m0157	-0.050	0.179	0.085	-0.089	0.159	0.104	-0.141	0.057	0.032
平均	-0.031	0.252	0.134	-0.101	0.164	0.108	-0.175	0.104	0.076
	$BETWEEN_{lp-lp}$								
a01m0037	-0.096	0.000	0.000						
a01m0086	-0.019	0.000	0.000						
a01m0088	-0.003	0.250	0.024						
a01m0124	-0.042	0.000	0.000						
a01m0157	-0.018	0.000	0.000						
平均	-0.036	0.050	0.005						

8.3 聴取実験

最も良かった組み合わせ実験結果 (表 10) で得られた要約について要約された講演音声を作成した。これを機械による要約とする。被験者 8 名により機械の要約と人間の要約結果 (man3) を聴き比べ、要点 (講演の内容がつかみやすかったかどうか) と聴きやすさ (自然な音声に聴こ

表 9 組み合わせパターンと各特徴の寄与度

特徴	書き起こし			音声認識結果		
	best1	best2	best3	best1	best2	best3
CUE_{all}	○ 0.2	○ 0.2	○ 0.2	○ 0.2	○ 0.2	○ 0.4
$PAUSE_{pre5}$			○ 0.2		○ 0.2	
$TAIL_{10}$	○ 0.4	○ 0.6	○ 0.6	○ 0.6	○ 0.6	○ 0.6
LEN						
TF	○ 0.2	○ 0.2	○ 0.2	○ 0.2	○ 0.2	○ 0.2
$PAUSE_B10_{pre5}$	○ 0.4	○ 0.2		○ 0.6		○ 0.6
$WORD_{prt}$	○ 0.4	○ 0.2	○ 0.4	○ 0.4	○ 0.6	○ 0.6
$\#LEN_{short}$	○	○	○	○	○	○
$\#BETWEEN_{lp-lp}$	○	○	○	○	○	○

表 10 組み合わせ実験結果

(a) 書き起こしを対象

講演音声	総文数	man3(文数)	system(文数)	一致文数	κ 値	適合率	F 値
a01m0037	184	75	61	46	0.490	0.754	0.676
a01m0086	313	77	103	44	0.289	0.427	0.489
a01m0088	281	78	93	50	0.405	0.538	0.585
a01m0124	185	54	61	35	0.433	0.574	0.609
a01m0157	333	90	111	64	0.482	0.577	0.637
平均	259.2	74.8	85.8	47.8	0.420	0.574	0.599

(b) 音声認識による自動書き起こしを対象

講演音声	総文数	man3(文数)	system(文数)	一致文数	κ 値	適合率	F 値
a01m0037	184	75	61	44	0.444	0.721	0.647
a01m0086	313	77	104	39	0.207	0.375	0.431
a01m0088	281	78	93	49	0.388	0.527	0.573
a01m0124	185	54	61	33	0.383	0.541	0.574
a01m0157	333	90	111	65	0.496	0.586	0.647
平均	259.2	74.8	86.0	46.0	0.384	0.550	0.574

表 11 韻律情報の寄与度が零の場合の組み合わせ実験結果

講演音声	書き起こし			音声認識結果		
	κ 値	適合率	F 値	κ 値	適合率	F 値
a01m0037	0.513	0.770	0.691	0.420	0.705	0.648
a01m0086	0.222	0.385	0.442	0.191	0.365	0.420
a01m0088	0.254	0.441	0.480	0.238	0.430	0.462
a01m0124	0.232	0.443	0.470	0.282	0.475	0.504
a01m0157	0.426	0.541	0.597	0.454	0.559	0.622
平均	0.329	0.516	0.536	0.317	0.507	0.531

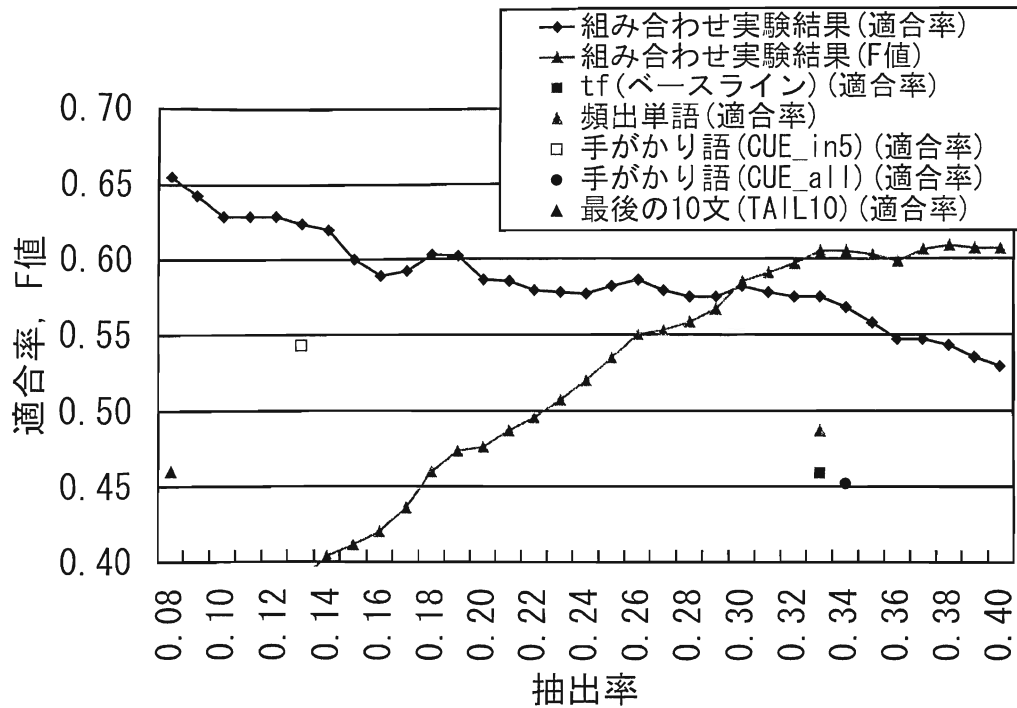


図 3 抽出率と適合率の関係

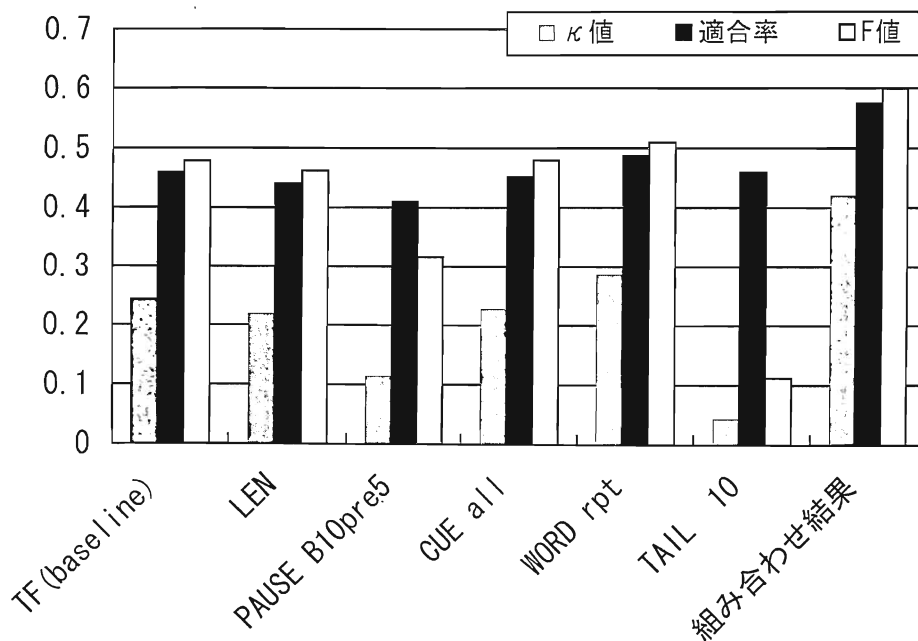


図 4 主な特徴の重要文抽出結果

えたかどうか)の2点について“人間の要約の方が良い”, “機械の要約の方が良い”, “どちらとも言えない”の3つの評価をした. この時, 各被験者には人間と機械のどちらの要約音声を聴いているという情報を与えていない. 一般に, 人間は後に聴いた方が良いと判断する傾向があり, そういった傾向をなくすため, はじめに機械の要約を聴く被験者数と, はじめに人間の要約を聴く被験者数を同じにした.

表 12 に聴取実験結果を示す. この結果より要点については機械の要約よりも人間の要約の方が良いという意見が過半数を占め, 人間の要約の方が優れている結果となった. この際, 聴きやすさの面については, 3項目ともほぼ同数の意見となり, 機械の要約が人間の要約とほぼ大差ないものとなったと考えられる.

表 12 聴取実験結果

(a) 要点

講演音声	人間の要約 > 機械の要約	どちらとも言えない	人間の要約 < 機械の要約
a01m0037	5	1	2
a01m0086	3	2	3
a01m0088	4	3	1
a01m0124	6	1	1
a01m0157	3	3	2
合計	21	10	9

(b) 聴きやすさ

講演音声	人間の要約 > 機械の要約	どちらとも言えない	人間の要約 < 機械の要約
a01m0037	3	3	2
a01m0086	1	4	3
a01m0088	3	2	3
a01m0124	5	1	2
a01m0157	1	5	2
合計	13	15	12

9 おわりに

本研究では, 人間による抽出要約の比較と, 表層的言語情報および韻律情報を用いた講演音声の自動的な重要文抽出による要約を試みた.

人間が行った抽出要約においては, 被験者間での κ 値は 0.286 とあまり高い値は得られなかった. これは対象とした講演が約 13 分間の研究報告ということもあり冗長な部分が少なく, また内容が高度に専門的でありどこが重要であるかの判断が被験者によりばらつきが大きいためと考えられる. しかし, NHK 番組「あすを読む」の要約に関しては被験者間の κ 値は 0.222

とさらに低かった。そこで、要約文集合の安定化を図るために、被験者 5 人中 3 人以上が重要文と認定した文を重要文とした要約 (man3) を基準とし、これと被験者の要約との κ 値は 0.407 となった。

次に、韻律情報と表層的言語情報を用いた重要文抽出の手法について示した。また、韻律情報と表層的言語情報を組み合わせることによって、より良い要約ができることが示された。最終的に複数の特徴を組み合わせることにより、人手による書き起こしデータにおいて κ 値で 0.420, F 値で 0.599 という良い結果が得られ、人間並みの重要文抽出ができた。また、講演音声の音声認識による自動書き起こしデータにおいても F 値が 0.574 と従来手法と比べて比較的良い結果を得ることができた。

今後は、韻律情報および表層的言語情報を用いた新たな重要文抽出の手法の検討や、学会講演以外の講演や講義に対する評価などを行いたい。

謝辞

音声要約の初期バージョンの開発にご協力いただいた本学大学院生の吉川裕規氏に深く感謝いたします。また、適切で有益なアドバイスをいただいた北岡教英講師に深く感謝いたします。

参考文献

- 赤松裕隆, 花井建豪, 甲斐充彦, 峯松信明, 中川聖一 (1998). “新聞・ニュース文をタスクとした大語彙連続音声認識システムの評価.” 情報処理学会第 57 回全国大会講演論文集, pp. 35–36.
- Carbonell, J., Geng, Y., and Goldstein, J. (1997). “Automated query-relevant summarization and diversity-based reranking.” In *IJCAL-97, AI and Digital Libraries*.
- Carletta, J., Isard, A., Kowtko, J. C., Doherty-Sneddon, G., and Anderson, A. H. (1997). “The Reliability of a Dialogue Structure Coding Scheme.” In *Computational Linguistics*, **23** (1), pp. 13–31.
- Edmundson, H. P. (1969). “New methods in automatic abstracting.” *Journal of the Association for Computing Machinery*, **16** (2), pp. 264–285.
- Fleiss, J. L. (1971). “Measuring Nominal Scale Agreement Among Many Rater.” In *Psychological Bulletin*, Vol. 76, pp. 378–382.
- 長谷川将宏, 秋田祐哉, 河原達也 (2001). “談話標識の抽出に基づいた講演音声の自動インデキシング.” 情報処理学会, SLP-36-6, pp. 35–43.
- Inderjeet Mani 著., 奥村学, 難波英嗣, 植田禎子訳 (2003). 自動要約. 共立出版.

- 井上章, 三上貴由, 山下洋一 (2002). “音声要約のための重要文検出における韻律パラメータの利用.” 日本音響学会秋季研究発表会, 2-9-20, pp. 101-102.
- 伊藤山彦, 松本賢司, 谷田泰郎, 柏岡秀紀, 田中英輝 (2001). “講演文を対象にした重要文抽出実験.” 話し言葉の科学と工学ワークショップ講演予稿集, pp. 157-164.
- 甲斐充彦, 廣瀬良太, 中川聖一 (1999). “単語 N-gram 言語モデルを用いた音声認識システムにおける未知語・冗長語の処理.” 情報処理学会論文誌, **40** (4), pp. 1385-1394.
- 笠原力弥, 山下洋一 (2001). “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, SLP-35-5, pp. 25-30.
- 河原達也, 李晃伸, 小林哲則, 武田一哉, 峯松信明, 嵯峨山茂樹, 伊藤克亘, 伊藤彰則, 山本幹雄, 山田篤, 宇津呂武仁, 鹿野清宏 (2000). “日本語ディクテーション基本ソフトウェア (99 年度版) の性能評価.” 情報処理学会研究報告, SLP-31-2, NL-137-7.
- 菊池智紀, 古井貞熙, 堀智織 (2002). “音声自動要約における重要文抽出の利用.” 日本音響学会秋季研究発表会, 2-9-18, pp. 97-98.
- 北出裕, 南条浩輝, 河原達也, 奥乃博 (2003). “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, SLP-46, pp. 7-12.
- 小林聡, 吉川裕規, 中川聖一 (2002). “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 音声言語情報処理研究会, 43-7 巻, pp. 41-46.
- 前川喜久男 (2001). “「日本語話し言葉コーパス」の構築.” 話し言葉の科学と工学ワークショップ講演予稿集, pp. 7-12.
- 松本裕治, 北内啓, 山下達雄, 平野善隆, 松田寛, 高岡一馬, 浅原正幸 (2003). 形態素解析システム『茶釜』version 2.3.3 使用説明書. 奈良先端科学技術大学院大学.
<http://chasen.naist.jp/hiki/ChaSen/>.
- 三上貴由, 井上章, 山下洋一 (2002). “ポーズで分割した発話単位の韻律パラメータと文重要度の相関.” 日本音響学会秋季研究発表会, 2-9-20, pp. 329-330.
- 三上真, 増山繁, 中川聖一 (1999). “ニュース番組における字幕生成のための文内短縮による要約.” 自然言語処理, **6** (6), pp. 65-81.
- 望月源 (2002). テキスト簡易要約器 Posum version 1.50.2 マニュアル. 北陸先端技術科学大学院大学.
<http://www.tufs.ac.jp/ts/personal/motizuki/software/posumcl/>.
- 長尾真, 宇津呂武仁, 嶋津明, 匂坂芳典, 井口征士, 片寄晴弘 (2000). 文字と音の情報処理. 岩波書店.
- 野畑周, 関根聡, 井佐原均 (2003). “異なるコーパスにおける重要文抽出の結果と素性の分析.” 自然言語処理, **10** (5), pp. 93-120.
- 野畑周, 関根聡, 井佐原均, R.Grishman (2002a). “自動獲得した言語パターンを用いた重要文抽

- 出システム.” 言語処理学会第8回年次大会発表論文集, pp. 539–542.
- 野畑周, 関根聡, 山元清貴, 井佐原均 (2002b). “話し言葉コーパスにおける文の切り分けと重要文抽出.” 第2回話し言葉ワークショップ, pp. 93–100.
- 奥村学, 難波英嗣 (1999). “テキスト自動要約に関する研究動向.” 自然言語処理, 6 (6), pp. 1–26.
- Salton, G. and Yang, C. S. (1973). “On the specification of term values in automatic indexing.” In *J. Documentation*, 29 (4), pp. 352–372.
- Siegel, S. and Castellan, N. (1988). *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw-Hill.
- 下岡和也, 河原達也, 奥乃博 (2002). “講演の書き起こしに対する統計的手法を用いた文体の整形.” 情報処理学会, SLP-41-3, pp. 17–24.
- 新中庸介, 菊池智紀, 古井貞熙 (2003a). “音声自動要約を用いた講演音声の速聞きシステム.” 日本音響学会春季研究発表会, Vol. 1, 2-4-5, pp. 67–68.
- 新中庸介, 菊池智紀, 岩野公司, 古井貞熙 (2003b). “音声自動要約を利用した講演速聞きシステムの検討.” 情報処理学会, SLP-46, pp. 13–18.
- TSC (2001). “Proceedings of the Second NTCIR Workshop on Research in Chinese & Japanese Text Retrieval and Text Summarization (NTCIR2).” In *National Institute of Informatics*.
- 宇津呂武仁, 小玉康広, 渡邊友裕, 西崎博光, 中川聖一 (2004). “機械学習を用いた複数の大語彙連続音声認識モデルの出力の混合.” 電子情報通信学会論文誌 D-II, Vol. J87-D2, No. 7, pp. 1428–1440.
- 宇津呂武仁, 西崎博光, 小玉康広, 中川聖一 (2003). “複数の大語彙連続音声認識モデルの出力の共通部分を用いた高信頼度部分の推定.” 電子情報通信学会論文誌 D-II, Vol. J86-D2, No. 7, pp. 974–987.
- Waibel, A., Bett, M., and Finke, M. (1998). “Tracking and summarizing meetings.” In *DARPA Broadcast News Workshop*.
- 山口優, 中川聖一 (2005). “講演・解説・ニュース音声の要約における表層的言語情報と韻律情報の検討.” 言語処理学会, 第11回年次大会, P1-2, pp. 173–176.
- 吉川裕規, 小林聡, 中川聖一 (2003). “表層的言語情報と韻律情報を用いた講演音声の要約と評価.” 日本音響学会講演論文集, 3-Q-35 巻, pp. 221–222.
- Zechner, K. (2002). “Automatic Summarization of Open-Domain Multiparty Dialogues in Diverse Genres.” In *ACL*, pp. 447–485.
- Zechner, K. (2003). “Spoken Language Condensation in the 21st Century.” In *Eurospeech 2003-GENEVA*, pp. 1989–1992.

略歴

小林 聡: 1991 年豊橋技術科学大学情報工学課程卒業. 1994 年同大学院工学研究科情報工学専攻修了. 1999 年静岡大学博士後期課程電子応用工学専攻満期退学. 1999 年豊橋技術科学大学知識情報工学系助手. 2000 年同大学情報処理センター助手. 2004 年島根大学総合情報処理センター助教授. 博士. コンピュータの教育利用の研究に従事.

山口 優: 2004 年豊橋技術科学大学情報工学課程卒業. 現在同大学院工学研究科情報工学専攻在学中.

中川 聖一: 1976 年京都大学大学院博士課程修了. 同年, 同大学情報工学科助手. 1980 年豊橋技術科学大学情報工学系講師. 1990 年同大学教授. カーネギーメロン大学客員研究員 (1985 - 1986)

音声情報処理, 自然言語処理, 人工知能の研究に従事. 工学博士. 1977 年電子通信学会論文賞, 1988 年度 IETE 最優秀論文賞, 2001 年電子情報通信学会論文賞受賞. 電子情報通信学会フェロー. 著書「確率モデルによる音声認識」(電子情報通信学会編), 「音声・聴覚と神経回路網モデル」(共著, オーム社), 「情報理論の基礎と応用」(近代科学社), 「パターン情報処理」(丸善) など.

(2004 年 9 月 28 日 受付)

(2005 年 2 月 9 日 再受付)

(2005 年 4 月 19 日 再々受付)

(2005 年 4 月 26 日 採録)