

論文 / 著書情報
Article / Book Information

論題(和文)	音声自動要約を用いた講演音声の速聞きシステム
Title(English)	
著者(和文)	新中庸介, 菊池智紀, 古井貞熙
Authors(English)	SADAOKI FURUI
出典(和文)	日本音響学会 2003年春季講演論文集, Vol. , No. 2-4-5, pp. 67-68
Citation(English)	, Vol. , No. 2-4-5, pp. 67-68
発行日 / Pub. date	2003, 3

1 はじめに

我々はこれまでで音声自動要約手法 [1][2][3] を提案してきた。本稿では、この手法による講演音声の要約結果を用いて要約音声を作成し、ユーザに提示する「速聞きシステム」を提案する。このシステムは、元の講演音声を用いて要約音声を作成するため話者の意図や感情といったテキストには含まれない情報を伝えることが可能である。また、要約結果をテキストで提示する場合には、要約結果に認識誤りが含まれているため、要約結果が読みづらく、ユーザに誤った内容を伝える可能性があるが、元の音声を提示することでこのような悪影響を取り除くことができる。

2 要約音声作成手法

提案する要約音声作成手法は、1) 音声認識し、その認識結果を単語や文を単位として削除・選択して、要約テキストを作成する「音声要約部」と、2) 得られた要約テキストに対応する音声区間を、元の音声からの切り出し、接続によって作成する「要約音声作成部」で構成される。以下では順に、その詳細を説明する。

2.1 音声要約部

音声認識は、文献 [3] と同じシステムを使用した。ただし、音響モデルについては日本語話し言葉コーパス (CSJ) 中の 94 時間 (455 講演) のデータを用い、混合数 16 の不特定話者音素文脈依存 HMM(2000 状態) を作成した。また、言語モデルについては約 1.4M 形態素を用いて語彙 26k の学習を行った。認識の際には、CSJ のラベルに基づき、講演音声を 500 ミリ秒以上の無音区間で区切り、その内 1 秒以内の音声は隣り合う音声と接続、20 秒以上のものはさらに分割した。以下で「文」と呼ぶものはこの 1 秒以上 20 秒未満の音声単位である。

要約テキストは、認識結果をもとに、文献 [1][2] の単語抽出手法と文献 [3] の重要文抽出手法を用いて作成する。単語抽出手法は、単語を単位として、重要文抽出手法は、文を単位として削除・選択を行い要約を作成する手法である。本研究では以下の 3 通りの単位を用いた要約を行った。

- 単語単位要約方式

単語抽出手法 [1][2] を用いて単語単位で要約を行う。重要部分を細かな単位で抽出することが可能であるが、短い単位の音声を多数接続することとなり、要約音声としては不連続性が生じやすく聞きづらくなる。なお、あらかじめフィラーを除いてから単語単位の要約を行っている。

- 文単位要約方式

重要文抽出手法 [3] を用いて文単位で要約を行う。音声としては連続性が保存されやすく、聞きやすいものになるが、単位が大きいため不要

な部分を含みやすい。なお、フィラーは最終的に除かれるため、「文単位要約」であっても、文内部のフィラー位置でテキストが切り離されている。

- フィラー間単位要約方式

認識結果を、文境界とフィラー位置で区切った単位を「フィラー間単位」とし、これを 1 つの文とみなし、重要文抽出手法を用いて要約を行う。単語単位と文単位の間隔的な削除・選択単位として導入したもので、「細やかな重要部分の選択」「連続性の保持による音声の聞きやすさ」という両者の利点を併せた効果が期待される。なお、フィラー自身は、区切った際に削除される。

2.2 要約音声作成部

音声要約部で得られた要約テキストは、各単位の削除・選択やフィラーの除去によって作成されているため、要約テキストに相当する音声を作成する際には、元の音声からの切り出しと接続を行う。その際には、認識結果の時間ラベルを利用して、要約テキストと切り出す音声区間の対応を取っている。

文の内部で音声を接続する場合には、2 つの音声の互いの音声のパワーの差から雑音が生じる可能性がある。これを防ぐため音声端のパワーを約 20 ミリ秒に渡って徐々に減衰させ、接続する。その際、接続部付近の音声の速度が速く、聞き取りにくくなるため、さらにポーズ区間を音声間に挿入する。ポーズ区間の長さは、経験的に 50~100 ミリ秒とした。このようにして作成された各文の要約音声を「要約音声文」と呼び、その音声区間に対応する要約テキストを「要約テキスト文」と呼ぶ。

要約音声文を作成した後、さらに接続を行って、講演全体の要約音声を作成する。この場合にも、文内部での音声接続と同様に、音声端のパワーを徐々に減衰させ、ポーズを挿入して接続する。ポーズの長さは短ポーズ (200 ミリ秒)、長ポーズ (700 ミリ秒) の 2 種類とした。具体的には、要約テキスト文の末尾の表記が「した・です・ます・せん・ある・ない・ね」となる要約音声文に後続の音声文を接続する際には長ポーズを、それ以外は短ポーズを挿入する。ただし、単語単位要約方式を用いる場合については、要約テキスト文に「体言止め」が頻出するため、このような要約音声文に後続の音声文を接続するときには長ポーズを挿入した。

3 評価実験

3.1 実験条件

CSJ 中の男声話者 3 名による講演音声 A01M0074 (M74), A01M0035 (M35), A05M0031 (M31) を要約率 50% で自動要約した。それぞれを認識した時の単語正解精度は約 70%, 55%, 70% であった。各講演に対し、単語単位・フィラー間単位・文単位の 3 種

* A quick listening system using automatic speech summarization techniques

表 1. 各講演に対する各方式の要約正解精度 (%)

	M74	M35	M31
単語単位	49.6	37.6	50.0
フィラー間単位	44.7	37.5	46.9
文単位	45.5	37.6	53.4

類の要約方式によって要約テキスト、要約音声を作成した。各講演音声の要約テキストについて、文献 [2] に従い正解要約文単語ネットワークに基づく要約正解精度を求めた。結果を表 1 に示す。正解要約文単語ネットワークは 9 人による正解要約文から作成した。

被験者には事前に講演の書き起こしを読んで内容を把握してもらってから、作成した 3 講演分の要約音声进行提示した。ただし、講演 M74 と M35 に関しては講演の予稿も書き起こしとあわせて事前に読んでもらった。各講演は、内容として意味のある 3 つの段落に分けられており、段落の順序は入れ替えず、それぞれの段落について 1 種類の要約方式で作成した要約音声被験者に提示される。1 つの講演で 3 種類の異なる要約方式で作成された要約音声被提示されることとし、その要約方式の並びは講演毎にランダムに入れ替えた。

評価基準は「聞きやすさ」「要約音声としてのふさわしさ」の 2 つで、「1: 非常に悪い, 2: 悪い, 3: 普通, 4: 良い, 5: 非常に良い」の 5 段階で評価を行った。被験者は 11 名である。

3.2 評価結果

「聞きやすさ」「要約音声としてのふさわしさ」の評価結果を図 1, 2 に示す。「聞きやすさ」について、単語単位要約方式は全ての講演において評価が一番低かった。講演 M74 と M35 については文単位とフィラー間単位要約方式がほぼ同等で、M31 については文単位の方が高かった。「要約音声としてのふさわしさ」についても、「聞きやすさ」の評価と同様な結果であった。全体として文単位要約が最も高い評価となった。

単語単位要約については、表 1 のように、要約正解精度が他の要約方式より高いにもかかわらず「要約音声としてのふさわしさ」の評価は最も低くなった。これは、短い音声区間の多数の接続による「聞きやすさ」の劣化が原因であると考えられる。

今回の実験では、文単位要約方式の評価が、「聞きやすさ」「要約音声としてのふさわしさ」とともに最も高かった。一方、フィラー間単位要約手法においては、意図した有効性を確認することができなかった。これは、要約率が 50% という条件では、文単位要約において不要部分が含まれることが少なかったこと、フィラーの頻度が多い話者では、フィラー間単位要約方式において、単語単位要約方式と同様の音声接続の頻発による「聞きやすさ」の劣化が生じることが原因であると考えられる。図 1 の M31 でフィラー間単位要約方式での性能が、文単位要約方式に大きく劣っているのは、後者の原因によるものと考えられる。表 2 には、各講演における各要約単位での、単位数を示している。M31 が他の 2 講演と比べ、フィラー間単位としたときに単位数が大きく増えていることがわかる。一方、M74, M35 の結果をみると、「聞きやすさ」における文単位要約からの大きな性能劣化が見られない。要約率をより小さく設

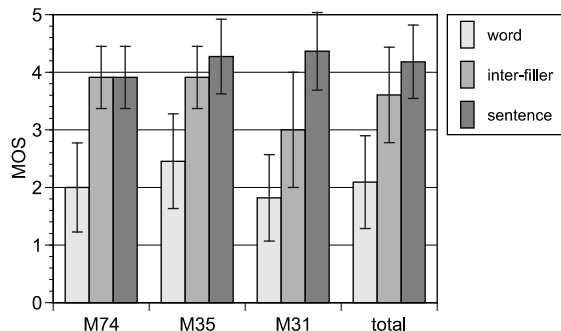


図 1. 「聞きやすさ」の評価結果

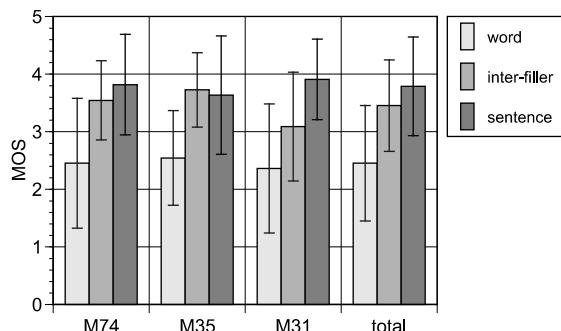


図 2. 「要約音声としてのふさわしさ」の評価結果

表 2. 各講演における単位数

	M74	M35	M31
単語単位	2,311	5,180	4,850
フィラー間単位	215	478	693
文単位	86	212	208

定した場合には、不要部分の影響によって文単位要約方式の「要約音声としてのふさわしさ」が劣化する恐れがあることから、このような講演に対しては、フィラー間単位要約方式の有効性が得られるものと期待される。

4 まとめ

本稿では、音声自動要約手法を用いた講演音声の「速聞きシステム」を提案した。単語単位・フィラー間単位・文単位の要約音声による評価実験を行い、50% の要約率の条件では、文単位要約方式が最も良好であったことを示した。

今後の課題としては、要約率を変化させた時の、3 種類の要約方式による要約音声の比較評価などがある。

参考文献

- [1] 堀智織, 古井貞照: “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).
- [2] 堀智織, 古井貞照: “講演音声の自動要約の試み,” 話し言葉の科学と工学ワークショップ講演予稿集, pp.165-171(2001-3).
- [3] 菊池智紀, 古井貞照, 堀智織: “重要文抽出と文圧縮による音声自動要約,” 電子情報通信学会 技術研究報告 NLC2002-81 / SP2002-158, pp.61-66(2002-12).