

キーワード抽出による映像コンテンツの理解支援とその心理的評価

岡本昌直 上地春奈 山本けい子 田村哲嗣 速水悟

岐阜大学 工学部

1. はじめに

近年、ニュース・解説などの独話をリアルタイムに理解することを支援するために字幕生成システムの開発が望まれている。字幕は聴覚障害者にとって重要な情報源であり、健常者にとっても、必要な情報かどうかすぐに分かるといった観点からも重要である。このため、商業放送においては、どのような字幕を入れれば短時間で送り手の意図を伝えることができるのか検討され、現在、NHK による字幕放送では、本文は部分要約されている。人間が 1 秒間に読み取れる文字数は 4 文字までといわれており、音声全てを字幕として表示すると、かえって読みにくくなるため、視聴者がわかりやすい字幕表示方法を検討する必要がある。

本研究では、キーワードを表示することにより、全文字幕においても要約された字幕と同程度の内容理解度を得ること、一瞬で内容を理解することを目的とする。また、一瞬で内容が把握でき、長時間観ていても疲れないことに重点をおき、心理的評価に基づいた字幕表示方法を検討した。

2. 先行研究

難波ら[1]はニュースに用いる文字テロップ画像を被験者に提示し、その読み取り時間を測定、その処理速度を算出した。それによると、漢字かな交じりの文では約 290 文字/分程度となった。アナウンサーの読み上げ速度は 300~400 文字/分程度といわれており、全文を字幕として表示すると十分に読み取れないおそれがある。

小峯ら[2]は、聴覚障害者にとって、リアルタイム全文字幕が有効か調査するために、全文字幕が音声と同じ速度で表示された場合と、従来の内容を要約して表示する方法との比較を行った。これによると、リアルタイム全文字幕を 280~350 文字/分程度の速さで表示すると、健常者の通常の試聴

状態の内容理解度とほぼ同程度まで改善されるという結果であった。また、5~7 割程度に要約された字幕とほぼ同程度の内容理解度が得られている。このことから、リアルタイム全文字幕は内容の理解に有効であることが示された。

これらの研究の他、音声認識を用いた音声要約やトピックモデルに関する研究[3,4,5,6]がある。

3. 字幕表示実験

本研究の提案手法で抽出したキーワードをどのように表示したら話題の理解を支援できるか、という心理的評価を行うために、被験者 20 名に 5 段階評価による実験を行った。スコアは評価項目の重み付き平均で算出した。

3-1. 表示方法に関する評価実験

字幕本文の見やすさ、分かりやすさといった評価に重点をおいて評価実験を行った。

表示方法、表示字幕範囲、字幕専用領域の 3 つの項目について、以下の 8 パターンで評価を行う。

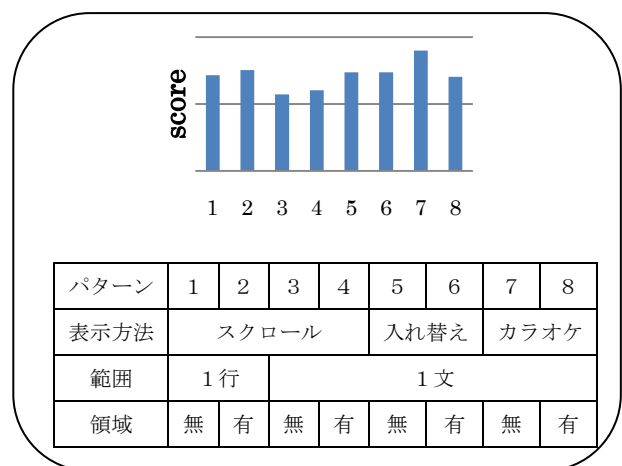


図 1: 表示方法の評価結果

表示方法のスクロールとは、字幕が右から左に流れ続けるものである。入れ替えは表示されている部分を読み終えたら次の字幕に切り替わるもので、現在の TV ニュースの字幕表示法と同じであ

る。カラオケ方式は字幕の読み上げている部分の色が変化する。表示字幕範囲1行とは、字幕を20文字1行として表示し、1文とは、句点までを表示する。結果は「カラオケ方式・1文・字幕専用領域無し」が最も良いと評価された。

3-2. キーワード表示のタイミングの評価

アンケートを行い、Aではニュース文の全体を複数の話題に分割し、各話題ごとに重要と考えられる単語を選択し、上位5単語を用いた。Bでは3文を1フレームとして同様に5単語を抽出した。

A:話題転換時、5単語表示

B:本文字幕が切り替わる度5単語表示

以上に加え、本文字幕表示方法（カラオケ方式/入れ替え方式）、音声の有無を組み合わせる評価実験を行った。結果は「A・入れ替え・音声有り」が最も良い評価を得た。

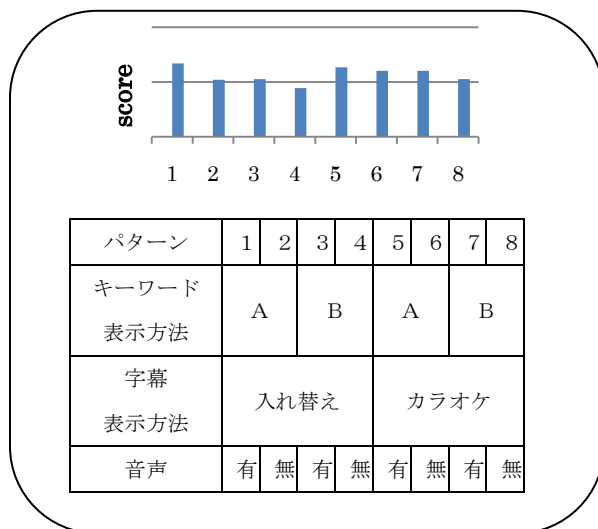


図 2:表示タイミングの評価結果

3-3. 字幕・キーワード表示位置の評価(1)

表示位置について以下の7パターンのうち、どれが最も見やすいか、話題の変化がわかりやすいかを評価した。表示位置は図3のとおり。パターン2, 3, 5, 6では表示位置(1)に字幕とキーワードの両方を配置した。結果は図4の7の組み合わせが最も良い評価を得た。

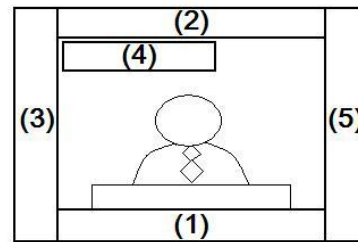


図 3:本文字幕・キーワードの表示位置

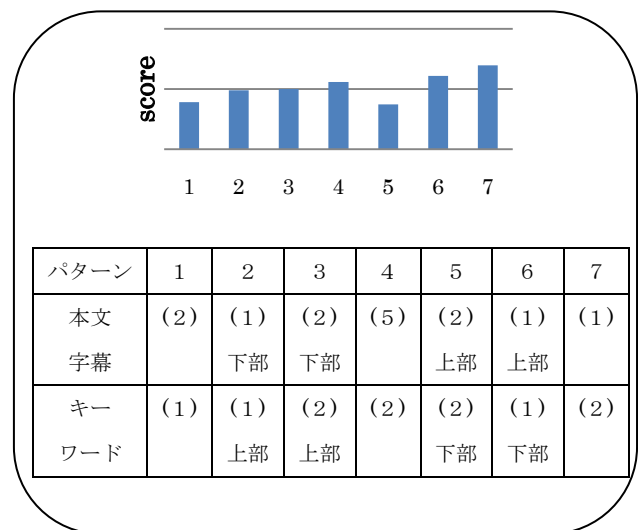


図 4:字幕表示位置実験パターン(1)とその評価結果

3-4. 字幕・キーワード表示位置の評価(2)

3-3の評価実験に補足として、新たに2つのパターンを加え、評価実験を行った。結果はパターン3の動画内にキーワードを埋め込むものが最も良い評価を得た。

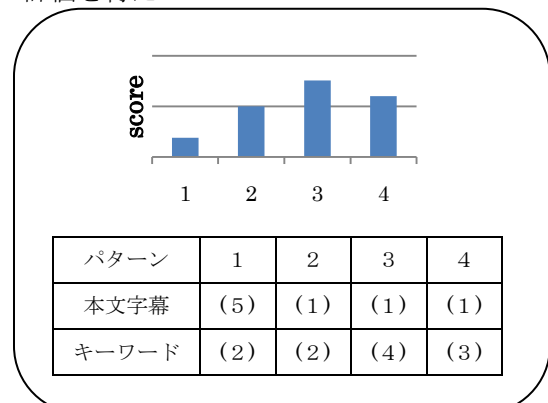


図 5:字幕表示位置実験パターン(2)とその評価結果

3-5. 結果・考察

本文字幕の表示方法は、一般的な入れ替え方式が好ましい。キーワードについては、表示するタイミングは、話題が変わったときが良い。これはあまり頻繁にキーワードが変化すると、本文字幕とキーワードの変化を短時間で把握しなければならず、疲れてしまうと考えられる。

表示位置に関しては、地震速報でみられる埋め込み型、もしくは台風速報などで見られる L 字型が良い評価を得た。日ごろ見慣れているために、このような結果になったと考えられる。また、本文字幕とキーワードが離れていると視点が定まらず集中できないといった心理的な影響も見られた。

4. キーワード自動抽出法

4-1. 最小二乗法による係数決定

形態素解析には形態素解析エンジン MeCab[7]、係り受け解析には日本語係り受け解析器 CaboCha[8]を使用した。

複合語を利用するために、wikipedia のページタイトルであり、2 単語以上からなる語を MeCab の辞書へ複合語として登録した。

以下の式のとおり単語重要度 \hat{y} を推定する。係数 ω は最小二乗法によって推定する。

$$\hat{y} = \sum \omega_i x_i + C \quad [1]$$

変数 x に用いたものと範囲は表 1 のとおりである。

表 1: x の値と範囲

	範囲
tf(単語出現頻度)	1~24
idf(逆頻出頻度)	1.1~5.6
tf×idf	1.2~69.3
接続名詞数	1~3
固有名詞か	0~1
係り受けされている節数	0~3

接続名詞数とは、その語がいくつの単語からなっているかを数えたものである。「日経平均株価」

の場合、「日経」「平均」「株価」の 3 単語からなるので、接続名詞数は 3 となる。

真値 y はアンケートによって決定した。このアンケートでは被験者 20 名、2008 年 10 月 16 日の NHK 時論公論の書き起こし文を使用した。重要と思う名詞を選択してもらい、その語を 3 段階でどの程度重要かを判断し、その重要度の文書内の総和をスコアとした。

4-2. 推定結果

x の組み合わせは以下の(i)~(iii)の 3 通りに対して係数 ω をそれぞれ推定した。結果をそれぞれ表 2-4 に示す。

(i) tf×idf, 接続名詞数, 係り受け数, 固有名詞かの判定

表 2:(i)における係数推定結果

	tf×idf	接続名詞数	係り受け数	固有名詞
係数	6.614	8.060	-2.078	13.08

(ii) tf, idf, 接続名詞数

表 3:(ii)における係数推定結果

	tf	idf	接続名詞数
係数	12.98	13.85	16.84

(iii) tf, idf, 係り受け数, 接続名詞数+固有名詞かの判定

表 4: (iii)における係数推定結果

	tf	idf	係り受け数	接続+固有
係数	12.89	13.31	-4.714	7.283

4-3. $y - \hat{y}$ 間の相関

相関を出すにあたり、すべての名詞を対象とした結果(A)と、対象をスコア(真値 y)が閾値を超えた名詞に限定した結果(B)との比較を行った。

(i)~(iii)における真値 y と推定結果 \hat{y} 間の相関係数は表 5 のとおりとなった。

表 5:(i)~(iii)における $y - \hat{y}$ 間の相関係数

	(i)	(ii)	(iii)
(A)	0.8173	0.6673	0.6752
(B)	0.9499	0.9556	0.9549

4-4. 考察

係数推定結果より、被験者が重要と判断する語は tf を重視していることが分かった。また、他に接続名詞数、つまり語の長さを見ていることも分かる。よって被験者は「出現回数が多く、長い語」を選ぶ傾向があった。また、表 5:(B)の結果をみると、閾値を超えた語、すなわち、被験者がより重要であると判断した語において、より高い相関がみられた。

5. キーワード抽出

キーワードとして表示される単語はどのようなものが好ましいかを調査するため実験を行った。

5-1. 実験条件

対象テキストを 3 文 1 フレーム、移動幅 1 文とし、各フレームより $tf \times idf$ が上位の単語を抜き出す。話題の転換、話題内容がどの程度把握できるかを調査することを狙いとした。その単語と同条件で被験者 20 名にてアンケートを行い、人間が重要と思う単語との比較を行った。

形態素解析には MeCab、係り受け解析には CaboCha、コーパスには、新聞社(5 社)の Web サイトの記事(2008 年 5 月~11 月に掲載された 96956 件)を使用した。実験には、2008 年 6 月 15 日正午の NHK ラジオニュースの書き起こし文を使用した。

5-2. 実験結果・考察

人手で抽出した単語と、 $tf \times idf$ によって抽出された単語には、高い相関はみられなかった。

$tf \times idf$ を用いた場合は、地名などの固有名詞が抜き出される一方、アンケート結果では内容を表す単語を抜き出す傾向がみられた。よって $tf \times idf$ を用いて得られたキーワードだけでは、十分に内容を伝えられるとは言い難い。また、単語では情

報量が少なく、複合語を表示させたほうが良いことがわかった。

6. まとめ

単語重要度の推定結果では、単語出現頻度と接続名詞数が重要視されるという結果が出たが、固有名詞、係り受けの情報が重要視されないとは考えにくい。また、時間経過に応じた単語重要度の変化も考慮に入れる必要があると考えられるので、推定方法についてさらなる考察が必要である。

字幕表示評価実験より、本文字幕部分は入れ替え方式で表示されることが好ましく、キーワードは頻繁に変化すると疲れ、本文字幕とキーワードの位置関係は、離れていないほうが集中しやすいという心理的評価を得た。

今後は被験者数を増やして実験を行う必要がある。抽出されたキーワードを用い、要約、トピック抽出等に利用できるかの検討も行う。

参考文献

- [1] 難波誠一:「画像情報と提示時間」NHK 技術研究所,35,1 (1983)
- [2] 小峰,星野,磯野,内田,岩鼻,榊原:「内容理解度によるリアルタイムニュース字幕の評価」映像メディア情報学会誌 Vol. 52,No.4(1998)
- [3] 堀,古井:「単語抽出による音声要約生成法とその評価」電子情報通信学会論文誌 D-II J-85-D-II NO. 2 pp.200-209 (2002)
- [4] 大野,松原,柏岡,稲垣:「同時的な独話音声要約に基づくリアルタイム字幕生成」情報処理学会研究報告 pp.51-56 2006-SLP-62-(10)(2006)
- [5] 門馬,沢村,三橋,原,白井:「聴覚障害者向けニュースの字幕提示方法に関する主観評価」映像メディア学会誌 Vol.54,NO.9,pp.1288-1297(2000)
- [6] 佐古,有木,岩田,渡部,堀:「話題の連続/不連続変化を考慮したトピックモデルに基づく音声認識」電子情報通信学会技術研究報告 SP-2008-88, pp.55-60(2008)
- [7] 形態素解析エンジン MeCab,
<http://mecab.sourceforge.net/>
- [8] 日本語係り受け解析器 CaboCha/南瓜,
<http://chasen.org/~taku/software/cabocha/>