

音声出力を目指した聞きやすいニュース読み上げの評価実験

佐藤 大[†], 増田 英孝[†], 中川裕志[‡]

東京電機大学工学部[†] 東京大学情報基盤センター[‡]

1 はじめに

近年、ブラウジング機能を持った携帯端末が普及している。一般的なコンテンツの出力は画面による表示である。しかし、運転中など目視で読むことができない場合には、音声で出力してくれる方が都合がよい。そこで、新聞記事サイトを対象に新たな出力方法として音声出力を提案する。携帯端末向け記事は小画面に表示することを目的としているため、体言止などの特殊な表現となっており、そのまま音声出力すると不自然な文章となる。そこで、パソコン向けに作成された新記事から重要文を抽出し、さらに音声出力に望ましい文章への要約と変換を行なう。

本研究では、まず文字数ではなく音声出力させる時間の目安をニュース番組で調査した。またいくつかのパターン記事データを音声出力させ読み上げに対する評価実験を行い、聞きやすい文章についての主観評価を行なった結果について報告する。

2 研究のアプローチ

新聞などの目で見える文章は、書き言葉と呼ばれている。それに対して会話は、話し言葉と呼ばれている。このことから、見ることに適している文章と、聞くことに適している文章があると考えられる。そこで音声出力の場合には、文章をそのまま音声として出力するのではなく、聞きやすい文章への変換をしてから出力することが望ましい。現在行われている要約・言い換えに関する研究は、見るためのものである。例えば、アナウンサーの発話からリアルタイムで字幕を作成する研究 [1] は行われているが筆者の知るところでは、音声出力向けに特化した文章要約・変換に関する研究は行われていない。そのために音声出力に適した要約・変換方法を実装するにあたり、調査を行う。

次に実装予定のシステムについて説明する。パソコン向けに作成されている新聞記事のタイトルを参考に、本文中から重要文を抽出し、音声出力に適した文への要約・変換を行い音声出力を行うシステムを考えている。

3 ニュース番組の調査

文章の要約は文字数を目安として行われている。例えば、携帯端末向けに新聞記事を要約する場合には、一画面におさまるように、50文字程度としている。本研究においては、音声出力を目指しているので、文章を読み上げる目安となる時間を決定したい。実際、携帯端末でニュースを音声で聞くのは短い時間を有効利用したい場合、例えば乗物の待ち合わせ時間など、時間圧の高い状況での利用が多いと考えられる。そこで、アナウンサーがニュースを読んでいる様子を見ることができ、一つのニューストピックスをとらえやすいので、ラジオではなくテレビで調査を行った。今回はNHKのTVニュース番組を参考にした。

3.1 調査の方法

ニュース番組においてアナウンサーが文章を読む時間を計測した。ニュース番組では一つのニューストピックスについて、はじめに短い時間で概要を述べている。本研究ではこの部分をニュースの冒頭部と定義する。その後、冒頭部より長い時間で、より詳しい内容を話している。ここをニュースの本体部と定義する。これは、新聞記事の見出しと本文の関係に似ている。見出しを見ればだいたいの内容を知ることができる。ニュースの冒頭部においても、その部分から本体部で話される内容を短時間で知ることができる。そこでニュースの冒頭部で文章が読まれる時間を、新聞記事を音声出力する際の目安とし、時間の計測を行った。

3.2 調査の結果

冒頭部の時間を計測した結果をまとめたものを図1に示す。調査ニュース数(トピックスの数)は232個である。15秒から17秒に集中していることがわかる。したがって新聞記事を音声出力向け文章に変換する際には15秒から17秒程度で読み上げ可能な長さとするのが自然な選択である。参考までに232ニュースのうち106ニュースについては本体部の時間を計測した

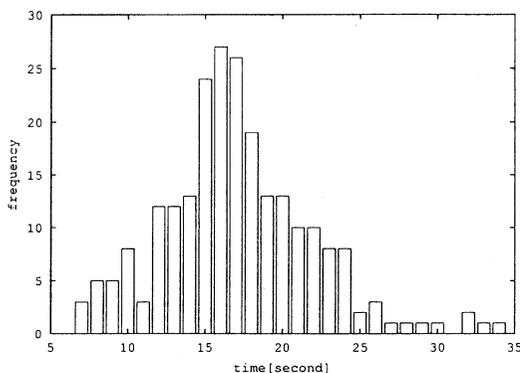


図 1: 冒頭部の時間

が平均 70 秒であった。

4 聞きやすい文章の調査

実際に新聞記事を音声出力させたものを被験者に聞いてもらい、聞きやすい文章(違和感なく聞ける文章)についての主観的評価を行った。使用したデータの詳細、評価実験の方法と結果を示す。

4.1 使用データについて

本研究では、携帯端末向け記事と Web 新聞記事の対応付けコーパス [2] を用いている。これは携帯端末向けに作成されている記事から、同じ内容の同じ Web 新聞記事を対応付けたものである。ここで携帯端末向け記事とはブラウジング機能を持った携帯端末での閲覧用に作成されている記事のことである(以下、携帯向け記事)。Web 新聞記事はパソコンでの閲覧用に作成されている記事である(以下、PC 向け記事)。そこから以下の 3 パターンの文章を生成した。

A: 携帯向け記事

例:〇月の月例経済報告で都の景気認識を示す基調判断を下方修正へ…東京都。
下方修正は 3 カ月ぶり。

B: 携帯向け記事の文に対応する PC 向け記事の文

例:東京都は△日、〇月の月例経済報告で都の景気認識を示す基調判断を下方修正する方針を固めた。

下方修正は 3 カ月ぶりとなる。

C: B を話し言葉に言い換えた記事

例:東京都は△日、〇月の月例経済報告で都の景気認識を示す基調判断を下方修正する方針を固めました。

下方修正は 3 カ月ぶりとなります。

A は携帯端末での表示を目的に作成されているので、50 文字程度である。また体言止めや「〜へ」「〜に」などの助詞で文が終わっている。さらに、「…東京都」のように文の最後に記事に関係する団体がくる独特な表現になっていたために、音声出力には向いてないと考えられる。また 82 記事について、読み上げた場合の平均は 11 秒なので、ニュースの冒頭部にくらべると短い。

そこで B として携帯向け記事の文に対応する文を PC 向け記事から自動抽出した。この場合は平均 26 秒である。対応文の抽出方法については、あとで説明する。

最後に C として、アナウンサーがニュースを読むことに習って、B の語尾を「です・ます調」言い換えを行い、年齢や前の単語を詳しく説明している「佐藤 (23)」のような () で囲まれた部分を削除した文章を用意した(なお、言い換えは現在、人手で行っている)。

サンプル 1 から始めてください。この場合はサンプル 1 からです。)

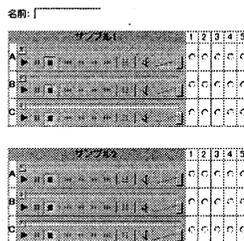


図 2: 評価用 web ページ

4.2 対応文の抽出方法

対応文の抽出は名詞の出現頻度によって行っている。携帯向け記事はほとんどが、2 文で構成されている(3 文の場合もある)。まず、携帯向け記事の 1 文目を茶筌 [3] を用いて名詞だけを取り出し、最も多くその名詞を含む PC 向け記事の文を対応文とする。次に 2 文目、3 文目と同様のプロセスを行うことにより、それぞれに対応する文を PC 向け記事の中から抽出するこ

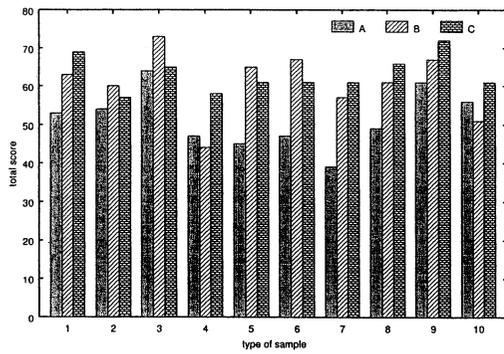


図 3: 合計点数

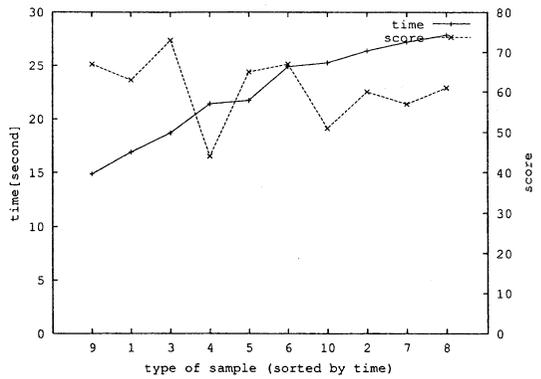


図 5: 対応文

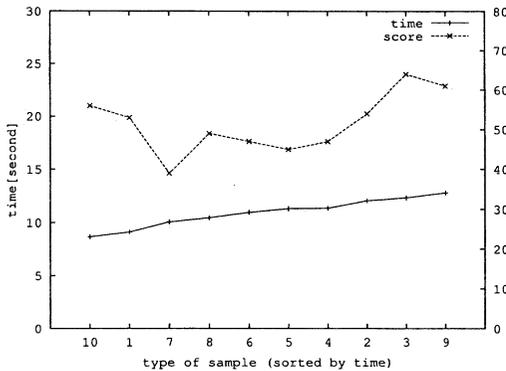


図 4: 携帯向け記事

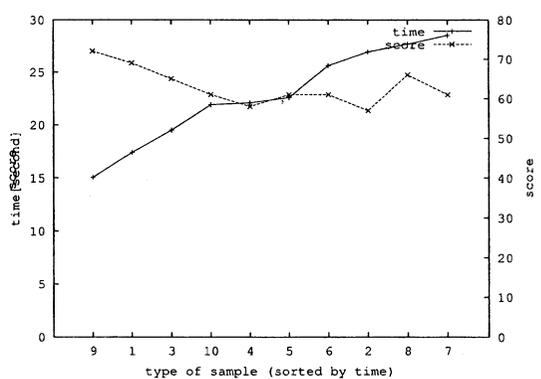


図 6: 言い換え

とが可能となる。ここで、例えば携帯向け記事の1文目に対してPC向け記事の3文目、携帯向け記事の2文目に対してPC向け記事の1文目が抽出された場合には、文章の流れがおかしくなることが考えられるため、PC向け記事の順序を基準とした文章にする。また携帯向け記事の1,2文目が共にPC向け記事の1文(同じ文)を対応文として抽出する場合もあるが、携帯向け記事の1,2文目の内容を含んでいる文を抽出しているので問題がない。

4.3 評価実験の方法

今回、文章の読み上げにはテキスト音声変換エンジン「SMART TALK Version3.0」沖電気を用いた[4]。ソフトの設定は、デフォルトの発話速度は普通(約6

文字/秒)で男性の声を用いている。このソフトで読み上げた文章をあらかじめ録音しておき、図2のような評価用webページを作成して実際に文章を聞いてもらい、以下のような点数をつけてもらう。最後には感想や意見を書いてもらう欄を設けている。点数付けは以下の指針で行ってもらった。

5. とても自然で聞きやすい。
4. 抵抗なく聞ける。
3. すこし不自然である。まあ、聞くことができる。
2. 不自然である。
1. とても、聞きにくい。

被験者には文章を見せずに音声だけを聞いてもらい、sampleにおけるパターンはランダムに配置しているの

で、被験者には、どのパターンの文章を聞いているのかは、わからないようになっていた。また評価を行う sample の順番もランダムである。

評価のために用意したデータは 10 種類の文章 (sample1~10) について、先ほどの 3 パターン (A,B,C) を作り合計で 30 となる。被験者は 19 人である。

4.4 実験の結果

はじめに、どのパターンが有効であるかを見るために被験者が評価した点数の合計を図 3 に示す。10 個のサンプルのうち合計得点が最も高くなっている数は 6 個で言い換えたパターン、4 個が対応文のパターンであった。総合的に見ると

言い換え > 対応文 > 携帯向け記事

の順番となった。やはり携帯向け記事は携帯端末の表示用に作成されており、特殊な表現のため点数が低い。対応文と言い換えについては、一般的な文章になっているので携帯向け記事に比べると点数が高い。さらに語尾を変換したことにより普段、聞いている文章に近いということで、対応文をそのまま音声出力させた時より聞きやすくなったと考えられる。

次に文章を読み上げる時間と得点に、どのような関係があるかを見るために携帯向け記事、対応文、言い換えに関してサンプルに対する時間と合計得点を示したものの (時間が短い順に並び替え) をそれぞれ図 4、図 5、図 6 に示す。まず携帯向け記事に関しては文字数が同じなので、音声出力による時間の差がほとんどなく、ここから時間と得点の関係は見る事ができない。対応文、言い換えに関しては、音声出力時間が長くなると合計得点が下がっていくことから長い文章は聞きにくいといえる。

また合計得点が高い sample1,9 は共に経済の記事である。経済の記事は「15 日 19 時現在、1 ドル〇〇円〇〇銭~〇〇銭」という形が決まっておりニュースでも同様な形であることから、聞きなれているそして、短い文書であるが情報の不足がないことも、得点が高くなった理由だと考えられる。

一般的に音声合成が望ましい場合とは 1, メッセージが単純。2, メッセージが短い。3, メッセージは、後で参照されることがない。など [5] といわれているが、被験者からは、「体言止め」は違和感があるという意見が多く、文章の語尾も重要であると考えられる。他には

「短かすぎると意味が理解できない」「聞き取りにくい単語がある」などの意見があった。

5 おわりに

本研究では、表示目的で作成されているニュース記事を音声出力するシステムの実装を目指している。そこで音声出力に適した文章を作るための参考としてニュース番組による調査と実際に音声出力した 3 パターンの文章を被験者に聞いてもらい主観的評価を行った。

まずニュース番組の調査から、15 秒から 17 秒程度で読み上げ可能な文章を音声出力に望ましい文章の長さとする。さらに、音声出力させた 3 パターンの文章を聞いてもらい聞きやすい文章の主観的な評価を行った結果、携帯向け記事をそのまま音声出力させた場合よりも対応文さらに、言い換えた文章が聞きやすい結果となった。また音声出力時間が長くなるにつれて、合計得点が下がるので、長い文章は音声出力には適していない。今回は対応付けコーパスを用いて携帯向け記事に対応する PC 向け記事の文を抽出したが、実装予定のシステムでは重要文を抽出して、音声出力に適した文への要約・言い換えを行なうので、重要文と対応文に関する評価を行なう予定である。

参考文献

- [1] 安藤彰男, 今井亨ほか: 音声認識を利用した放送用ニュース字幕制作システム, 電子情報通信学会論文誌, Vol. 84-D-II, pp. 877-887 (2001).
- [2] 大森岳史, 増田英孝, 中川裕志: Web 新聞記事の要約とその携帯端末向け記事による評価, 情報処理学会自然言語処理研究会, No. 153, pp. 1-8 (2003).
- [3] 松本裕治, 北内啓, 平野善隆, 松田寛: 形態素解析システム「茶茎」version 2.2.9 使用説明書, 奈良先端科学技術大学院大学松本研究室 (2002).
- [4] 沖電気工業株式会社: テキスト音声変換エンジン「SMART TALK Version3.0」(2000).
- [5] Shneiderman, B.: ユーザーインターフェースの設計, 6.4.4 音声合成, 日経 BP 社, 第 2 版 (1993).