

余剰語を用いた聞き取りやすい文章読み上げシステムの開発

○ 高野 文徳

藤波 努

Fuminori TAKANO

Tsutomu FUJINAMI

北陸先端科学技術大学院大学 知識科学研究科

School of Knowledge Science, JAIST, HOKURIKU

{ftakano,fuji}@jaist.ac.jp

概要

音声合成技術は誰にでも利用できるものとして普及してきている。また、合成音声の評価基準についてもある程度の指針 [粕谷 92][粕谷 93] が考えられている。また、間投詞の役割についての研究 [伊藤 99] もなされている。

本研究では、市販されている音声合成システム (IBM 社製 ProTalker97) を用いて単調な印象を与えない Text-to-Speech システムの構築について報告する。本システムでは、特に間投詞、感動詞などの余剰語を用いた構築する。余剰語を用いることで、注意が向くか、聞きとりやすくなるか、という効果について、余剰語を用いていない場合との比較、空白となっている場合との比較を行なう。また、音声合成システムにおける効果的な余剰語の付加について考察する。

1 はじめに

1.1 背景

音声技術は一般生活に浸透してきている。特に音声合成技術はマン・マシン・インタフェースとしてさまざまな場所で実用されている。また、実用に伴ない、音声合成の評価についての研究も様々な形で存在する。音声合成の評価方法についての基軸として、了解性、自然性、適合性からの考察が必要であることがあげられる [粕谷 92][粕谷 93]。

了解性、自然性、適合性のそれぞれは、

了解性 言語情報を正確に伝えているか

適合性 平均ピッチ周波数、変化幅、速度などの場面による適合

自然性 どこでどれだけ自然音声と異なるか

などの具体的な観点からの考察が考えられる。

また、間投詞の利用、評価の研究として、講演、対話システムにおける間投詞の種類、無音区間に変更したときの違和感、効果についての研究 [伊藤 99] があげられる。この研究では、あらかじめ録音された音声を加

工、もしくはあらかじめ録音してある PCM 音声を利用している。

容易に入手できる音声合成システムを用いて Text-to-Speech システムを構築する場合に問題になる点は、合成音声の品質に満足できるか、という点である。この不満を解消するために、あらかじめ録音してある音声を利用することが考えられる。しかし、大量の音声データが必要となってくる。そこで、いかに少ない操作で (合成システムのもつ機能を使って) 使用者の負荷を減らすことができるか、という点に着目した。

1.2 目的

本研究では、入手が容易な一般的な音声合成システムを用いて Text-to-Speech システムを構築する。そして、システムに余剰語を挿入することによって聞きとりやすさ、注意を向かせることができるかということを中心に余剰語の効果を調べる。また、より効果的な余剰語の使い方について考察する。

本論文において、余剰語とは間投詞、感動詞の類で意味をもたない語と定義する。(余剰語の例については表 2 を参照のこと)

1.3 論文構成

はじめに、余剰語を挿入していく指針を得るために、実際の対話をもとに余剰語の頻度、種類を抽出する (§2). 次に、Text-to-Speech を構成し、余剰語を挿入した場合、しない場合、空白をかわりとした場合を聞きとり調査し、まとめる (§3). 最後に結論 (§4) と今後の課題について述べる (§5).

2 対話コーパスの分析

2.1 分析の目的

余剰語がどのような頻度で出現しているか、新情報処理開発機構 (RWC) の対話コーパス (海外旅行の計画)¹をもとに分析した.

余剰語がどのような割合で出現しているのか、次に発話しようとしている単語の品詞によって出現頻度が違うのか、どのような種類のものが多いかを調べる.

2.2 分析の方法

形態素解析をし、意味をもたないであろう語を選別した. そのあと、目的に則し、文頭での出現の割合、品詞ごとの出現の割合、種類ごとの割合についてまとめた.

今回は、形態素解析のソフトウェアとして茶筌²を用いた. 辞書等は附属のものを使い、変更していない.

2.3 分析の結果

発話文は 1182 文であった. そのうち、余剰語は 1130 語存在した. 文頭に存在したものは、413 語 (36.5%) であった. また、品詞による分類に関しては、表 1 からわかるように、名詞の前に余剰語が発生している割合が高い (51.1%).

また、余剰語のあとに余剰語がある場合が 10.5% であることから、繰り返して発話されることも少なくないことがわかる.

種類では、表 2 からわかるように、「あー」「あの一」

表 1: 余剰語のあとにきている品詞

品詞		(%)
名詞	577	51.1
副詞	125	11.1
感動詞	72	6.4
動詞	38	3.4
接頭詞	29	2.6
連体詞	29	2.6
形容詞	21	1.9
接続詞	18	1.6
助動詞	14	1.2
助詞	9	0.8
未知語	32	2.8
余剰語	119	10.5
文末	47	4.2

表 2: 余剰語の種類

		%
あの一	195	17.3
え一	153	13.5
あの	136	12.0
あ	107	9.5
そうですね	74	6.5
あー	53	4.7
あっ	43	3.8
え	43	3.8
ん一	37	3.3
ま	36	3.2
え一と	27	2.4
ん	23	2.0
う	16	1.4
え一とと	16	1.4
えっ	15	1.3
は一	15	1.3

など「あ」からはじまるものが多い. これは、解析したコーパスが対話であるため、これから発話しようとしている内容が前に発話している相手に答えるかたちの場合が多いからである. 今回は Text-to-Speech システムを目的としているので、対話形式ではない. なので、受け答える的なものは無いと考えられる. よって、上記の結果については、参考程度にしかならないであろう.

3 評価実験

3.1 実験環境

特定のメールアドレスにアナウンスしたい内容のメールを送信すると、合成音声によってラジオからメール

¹<http://www.rwcp.or.jp/vswg/rwcdb/speech/>

²使用したバージョンは、2.0b6(Linux, Solaris) と 2.0(Windows98, Linux) である.

のサブジェクトが読み上げられる、という簡単なシステムを構築した。

今回はプロトタイプとして、朝日新聞社の asahi.com³ のニュースを利用してニュースの見出しを読み上げるシステムを構築した。読み上げのための音声合成に IBM 社製 ProTalker97 を用いた。

3.2 余剰語挿入アルゴリズム

本システムでは、余剰語は分析結果 (§2.3 参照) に基づいて挿入している。しかし、分析したコーパスは人間同士の対話なので、変更している。

変更するにあたり、分析結果 (§2.3) で得られた知見を忠実に実装したシステムを試作した。このシステムを 3 人に聞いてもらい、試作したシステムの改善点を明らかにした。具体的な変更点は、以下の 3 点である。

1. 文頭以外の余剰語の繰り返しをしない

文中において、余剰語を繰り返すとうっとうしさがあります。との報告をうけた。さらに、同じ語を繰り返してしまう場合 (「えー、えー」など) ともストレスを感じる、という感想だった。

2. 使用する余剰語の種類の選択

「あの一」「あー」など「あ」からはじまるものは、なにかの指示語のように聞こえるため削除した。「そうですね」は、対話の例では意味をもたないときもあるが、一方的な場合はなにが「そう」なのか意味不明、との意見があった。また、「んー」は、「うー」と聞こえ、うめいていて気持ちが悪い、とのことだったので、削除した。同様の理由で、「う」も削除した。「ま」は、一方的なシステムにおいては不自然すぎるということで削除した。

3. 名詞の繰り返しの場合、先頭の名詞のみを余剰語の挿入の対象とする

名詞が繰り返して一つとなっているもの (姓名など) の途中で余剰語が入ることは、人間ではありえないことなので、繰り返しの間は挿入されないようにした。

³<http://www.asahi.com/>

3.3 実験方法

合成音声を漫然と聞き流してしまうことを防ぐため、合成音声を聞きながら書取りをしてもらった。計 3 種類の実験を行なった。3 種の内訳は、(1) 余剰語有り (2) 余剰語のかわりに 4 秒程度の空白 (3) 余剰語・空白無しである。実施する順番によって 6 つのグループにわけた。そして毎作業後、感想を聴取した。疲労を考慮して、1 日 1 種とし、また前の回の印象が薄れないよう、連続した日で実験をおこなった。読み上げの速度は、1 秒間にひらがなにして約 5 文字 (設定可能中最低速) である。

グループによって聞いてもらう組合せは、以下のとおりである。

	グループ No.					
	1	2	3	4	5	6
実験 1	×	●	○	×	●	○
実験 2	●	○	×	○	×	●
実験 3	○	×	●	●	○	×

× : 余剰語、空白無し

● : 余剰語あり

○ : 余剰語のかわりに空白

実験は各グループ 2 名ずつ、合計 12 名の被験者を対象とした。

また、被験者の殆どは、システムで使用した合成音声に対して初めて接するので、合成音声に対する慣れ [粕谷 92] を考慮して、1 回練習してもらった。練習でのデータは、2000 年 1 月 10 日～12 日までの asahi.com のニュースの見出しをそれぞれ 1/3 ずつ書き取りをしてもらった。

今回使用したデータは、できるかぎり既知の情報ではなく、また各人の知識による差を減らすために、実験を行なうそれぞれの日 (2000 年 1 月 18 日、19 日、20 日) の asahi.com の 12:30 の時点のニュースの見出しを用いた。実験はそれぞれの日の午後に行なった。また、余剰語は擬似乱数を用いて挿入しているが、実行ごとに挿入するところが変化しないよう、種を固定して実験を行なった。また、それぞれの見出しの間には、約 1 秒の空白がある。

3.4 実験結果

見出しの形態素解析をした結果、読み上げた単語数は、実験 1 707 語、実験 2 739 語、実験 3 736 語であっ

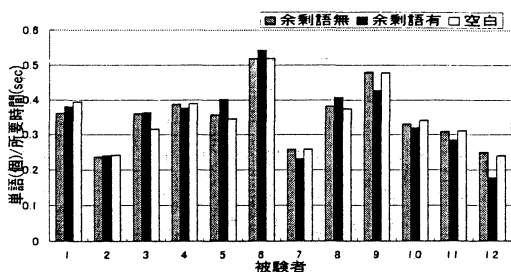


図 1: 一秒あたりの単語数

た。読み上げた文の数は、それぞれ 67 文で、所要時間は、余剰語の無い場合が約 8 分、余剰語のある場合が約 9 分、余剰語のかわりに空白を用いた場合が約 10 分であった。単語の一部でも記述が認められたら、数えあげる対象とみなして単語数を数えたところ、個人による差が大きい、およそ 100~300 程度の単語を書きとることができていた。図 1 が時間 (1 秒) あたりの単語数である。

図 1 からわかるように、空白の場合と無しの場合との差異は小さい。余剰語が挿入されたときは、邪魔と感じた人と、あってもあまり苦にはならない人によって結果の逆転が見られた。

4 結論

実験から明らかになったのは以下の 4 点である。

1. 余剰語の挿入により注意が向く

余剰語がある場合に書く注意が次の単語へと向くことが多くなるため、書ける単語数が増加したと考えられる。ただし、不自然だ、不愉快と感じた人がいること、邪魔と感じて逆の効果を産んでしまっている場合もあることは否めない。

2. 文の途中では余剰語は入らない方が良い

余剰語があると文の切れ目と感じて、次の文へ進んでいくことが多々見うけられた。つまり、余剰語は文の切れ目の指標として使用したほうがより効果的であろう。

3. 余剰語には特別な機能がある

余剰語のかわりに空白を入れた場合と、何も入っていない場合との差が無いといえる結果を得てい

る。このことから、余剰語が空白と同じように働いているとは考えにくい。何らかの固有の効果があると考えられる。

4. 空白の場合にシステムが停止していると感じる人がいる

これは、[伊藤 99] においても言及されていることの再確認となった。

5 今後の課題

使用した音声合成システムにおいて設定できるパラメータが速度、ピッチの高低、抑揚の強弱、性別など大雑把なものだけであったので、余剰語のイントネーションが不自然であった。より自然なイントネーションにただけでも、結果がかわってくると考えられる。また、間のとりかたに関しても自由度のあるものならば余剰語が効果的に挿入できるであろう。

また、課題に関しては、意味がとらえられているかどうか、つまり了解性の向上につながるかどうかはかることのできる課題を設定し、既存のシステムに余剰語を付加した場合の了解性に関する部分の研究をしていきたい。

謝辞

本学の石崎雅人助教授には、RWC 対話コーパスデータを示していただき、また貴重な助言をいただいた。感謝いたします。

参考文献

- [伊藤 99] 伊藤, 峯松, 中川: 間投詞の働きの分析とシステム応答生成における間投詞の利用と評価, 日本音響学会誌, Vol. 55, No. 5, pp. 333-342, 1999.
- [粕谷 92] 粕谷: 音声合成技術の評価, 日本音響学会誌, Vol. 48, No. 1, pp. 46-51, 1992.
- [粕谷 93] 粕谷: 合成音声の品質評価法, 日本音響学会誌, Vol. 49, No. 12, pp. 866-870, 1993.