朗読における文書の構造と間の関係に関する考察

青山 明史
東京理科大学*

田中 (石井) 久美子
東京大学†

武田 正之
東京理科大学*

1　はじめに

今日、音声合成器は我々の身近なところで使われるようになってきた。例えば、会話をするロボットの研究開発に使われる以外にも、日常生活において家電の状態を知らせたり、外国語を学ぶ人が発音を知るために使われたりする。この音声合成器の技術は誤った場所に間を入れず、平均的抑揚の韻律を用いて文書を読むことで、ある意味実用に耐えうる精度にまで確立したといえるだろう。

しかし音声合成器は文脈をふまえた上で自然な抑揚で文書を読み上げることができるとは未だいいがたい。つまり、表現豊かに文書を読み上げるという点においては研究途中にある。とはいえ、NassとReeves[1]は文書を読み上げる際に、音声合成器が抑揚と文脈の整合性は聴き手にとって重要であることを心理実験により示した。例えば、悲しい内容の文書を音声合成器に入力し、それを陽気な音声で読み上げると、聴き手は不愉快に感じる。

技術的に表現豊かな音声を生成することは、さまざまな音声合成器に用意されている抑揚、強勢、間を制御するためのコマンドを入力文書に埋め込むことで、可能である。そのためには、人間の音声の抑揚、強勢、間のとり方の法則を調査する必要がある。

この人間の音声の抑揚、強勢、間のとり方の研究はまだ始まったばかりである。韻律について、Pan[2]は医者と患者の診断の会話において、文脈が抑揚に影響を及ぼすことを研究した。Fry[3]は日本語の会話で、新しい主題が現れると、そこにアクセントが置かれることを示した。間についての研究は音声認識や自然言語理解が主であり、間をとる場所や長さについてはあまり研究されていない。宮崎ら[4]は韻律を用いた間のとり方の研究を行っているが、構文や意味を反映した間のとり方の必要性が述べられている。

杉藤[5]は、朗読の教本において、間が聴き手の印象に影響を及ぼす事を示した。例えば、話し方が早口であっても、直後の間を長くとれば早口とは聞こえないなどである。

我々は杉藤の研究成果をもとに、朗読データの間の解析を行った結果を報告する。これは将来、音声合成器を用いて間の場所と長さをより適切にとることにつながる。

2　仮説

杉藤は、人間が文書を読み上げる時に息継ぎのほか、読み上げている内容を聴き手によく伝えるために、間をとる場所や長さを調節していることを述べている。そのため、間をとる場所は構文、意味、強調と関係があることを示している。このほか、文書中の文節、文、段落という要素に対して、大きな要素のあいだほど長い間がとられる傾向にあるという仮説を立てている。そこで我々は、要素の大きさを「要素内の平均文字数」と定義し、大きさの異なる要素のあいだの間の長さの比較をおこなった。

さらに要素の大きさだけでなく、それぞれの要素と要素の関係を定め、その関係がどのように間の長さに影響を及ぼすのか調査するため、以下のような仮説を立てた。

仮説：文書内の2つの要素の関係が近いほど、そのあいだにはより短い間をとる

ここで、要素の関係を以下のように定義する。

【文節】　それぞれの文節の係り受けの距離
【文】　それぞれの文の主語が同じかどうか
【段落】　それぞれの段落の単語頻度ベクトルの内積

以上のほかにも要素の関係を定義する方法はあったが、上記の関係は自動的に計算可能であることから、これらを選択した。

3　データと間の抽出

3.1　データ

実験に用いるデータとしては、会話やニュースのアナウンスなどさまざまなものが考えられるが、本研究においては以下の4つの制約を設けた。

* 東京理科大学 大学院 理工学研究科 情報科学専攻
† 東京大学 大学院 情報学環

表 1: 解析に用いた朗読データ

タイトル	河童	坊ちゃん	蜘蛛の糸
朗読者	橋爪 功	風間 杜夫	加藤 武
長さ	31分57秒	42分49秒	11分38秒
単語数	5016	8379	1644

1. 音声データに、感情を込めたり、表現豊かに発話している部分があること。
2. 文書のみが発話に影響を及ぼし、会話での相槌やジェスチャーなどの要因が影響を及ぼさないこと。
3. BGMやノイズが含まれていないデータであること。これは音声データを3.2節で述べる信号処理にかけ、間を自動的に検出するためである。
4. 読み上げている文書の電子文書を取得可能であること。これは電子文書を音声合成器に入力し、人間と音声合成器の音声データを比較するためである。音声合成器は「おしゃべりメイト」[6]と「Microsoft Agent」[7]を使用した。

研究で用いた以上の制約を満足する3つのデータ[8]を表1に示す。これらは朗読を職業とするの男性が単独で朗読を行っているものである。

3.2 間の抽出

まず朗読データが読み上げている文書を、2章で示した要素に分ける作業を行った。段落、文は文書の句点と改行により抽出可能である。次にKNP[9]を用いて文節とその修飾関係を取得した。

一方、朗読データ内の間の抽出は信号処理により行った。我々が利用した朗読データは、録音スタジオ内で録音されたデータであるので、朗読者以外の音声やノイズが少ない。つまり間の場所は最大出力が0に近い所として特定可能である。従って、最大出力がある閾値以下のすべての場所を間として抽出した。

こうして文書の要素の区切れの場所と朗読データの間の場所を抽出した。この2つを2段階に分けて対応付けを行った。まず最初に動的計画法を用いて文、段落の対応付けを行った。次に各文において文節単位での対応付けを手動で行った。

4 解析結果

4.1 要素の大きさと間の長さ

図1は文節、文、段落の各要素の大きさと、間の長さの平均の関係を表す図である。横軸は要素の大きさの平均文字数の対数であり、縦軸はその後につづく間の平均の長さである。文節、文、段落の平均文字数は各々3.95、31.79、343.95文字であった。また各要素の標準偏差を表2に示す。

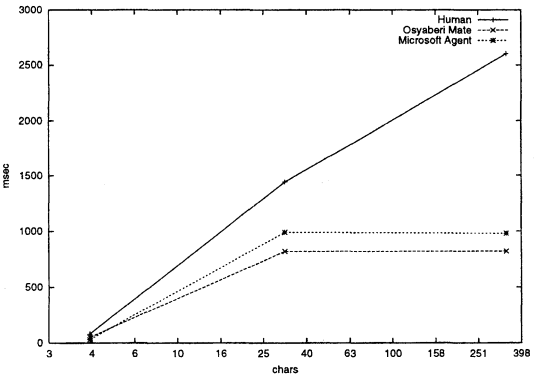

図 1: 要素の大きさと間の長さの平均

表 2: 間の長さの標準偏差

	人間	おしゃべりメイト	MS Agent
文節	200msec	70msec	80msec
文	640msec	30msec	120msec
段落	1540msec	30msec	70msec

図1から、人間の朗読の場合はより大きな構造の後に長い間が観察される。すなわち

(文節の間の長さ) < (文の間の長さ) < (段落の間の長さ)

という関係が導かれ、一方、音声合成器では、長さ、標準偏差ともに

(文節の間の長さ) < (文の間の長さ) ≈ (段落の間の長さ)

という関係となっている。

また、表2の人間と音声合成器の各要素の標準偏差に注目すると、音声合成器の標準偏差は人間にくらべて小さいことが観察される。つまり人間の間にはゆれがあるのに対し、音声合成器では全くないことが観察される。

そこで、4.2節では、人間の朗読の間のゆれと要素の関係を解析する。

4.2 要素間の関係と間の長さ

我々は何が間の長さに影響を与えるのかに注目した。これを2章で述べた仮説を検証することで考えていく。

4.2.1 文節

2章において2つの文節の関係は、2つの文節の係り受けの距離を用いて表すものと定めた。例えば以下のような構文木をもった文の文節の距離を考える。

例文：チャックは僕を小ぎれいなベッドの上へ寝かせました。

文節の係り受けの構造はKNPを用いて調べることができ、例文の文節の係り受けの構造は図2のようにな

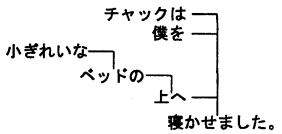

図 2: 係り受けの構造

る。これは、「チャックは」「僕を」「上へ」は「寝かせました」に係り、「小ぎれいな」は「ベッドの」に、「ベッドの」は「上へ」に係ることを表してる。この時、例えば「チャックは」と「寝かせました」の関係は、「チャックは」の5文節先に「寝かせました」が存在するので、距離は5とした。これは最初の文節と修飾される文節間の文節数に一致する。また、「チャックは」の文節を距離5の文節と呼ぶことにする。例文の文節はすべて以下のように分けられる。

距離5の文節 「チャックは」
距離4の文節 「僕を」
距離1の文節 「小ぎれいな」「ベッドへ」
「寝かせました」

距離 n の文節と間の関係を調査するために、距離 n の文節に対する間の平均長を計算した。文書中に長い文の頻度が少なく、しかも長い文から文節の正しい構文木を生成する可能性が低いため、生成された構文木はすべて手動で確認し、正しい構文木の文節に対してだけ計算するようにした。従って $n > 4$ の文節は30に満たない。

距離 n の文節 (横軸) と間の長さ (縦軸) の関係を図3に示す。まず、全体的に人間は n が増加するにつれ間の長さが長くなっているが、音声合成器ではほぼ一定であることが観察される。

次に人間の間の長さは $n \leq 4$ では単調増加であるが、$n > 4$ でその傾向が崩れてしまうことが観察される。これには以下の2つの理由が考えられる。

- $n > 4$ 以上の文節の標本数が少ないので、十分な平均値が得られなかった。
- 人間が文書を読み進めながら文の構造を理解する限界の距離が4であった。

また図3に示した平均よりも長い間をとる文節について調べると、それらの多くは関連する文節を強調していることが観察された。

4.2.2 文

連続する2文の関係を

- 2文の主語が一致する → 2文の関係が近い
- 2文の主語が一致しない→ 2文の関係が遠い

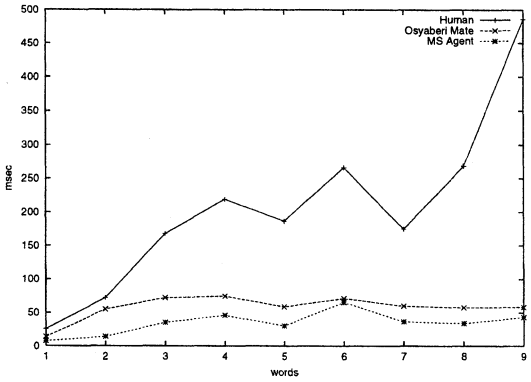

図 3: 距離 n の文節と間の関係

と本稿では定義する。これを用い文の関係が間にどよのうな影響を及ぼしているのかを調査した。

まず、2文のあいだの間が平均よりも短いところをランダムに100対選び、その主語の一致度を計算した。この100カ所の間の長さの平均は775.9msecであった(図1よりすべての文の後の間の長さの平均は1444.1msecである)。その主語の一致度は60%であった。また40%の2文についてより詳しく調査した。すると、そのほとんどが主語が相互に関係のあるシーンであった。以上より、関係が近い2文のあいだの間は平均よりも短くなる傾向にあるといえる。例えば以下の2文である[1]。

1. そのうちに僕は飛び立つが早いか、岩の上の河童へおどりかかりました。
2. 同時に河童も逃げ出しました。

この2文のあいだの間は590msecである。文1の主語「僕」と文2の主語「河童」が格闘している関係のシーンである。

次に、2文のあいだの間が平均よりも長いところをランダムに100対選び、その主語の一致度を計算した。100カ所の間の長さの平均は2344.5msecであった。すると、26%しか2文の主語が一致しいていなかった。つまり関係が遠い2文のあいだの間は平均よりも長くなることがいえる。また主語は一致するがあいだの間を平均より長くとった26%について調べると、強調のために間を長くとっていることが観察された。例えば、前の文が主語について描写し、その後に長めの間をとることで、つづく文の主語を強調することが挙げられる。

以上をまとめると、2つの文の関係が近ければそのあいだの間は短くなる傾向にあることが観察された。特

[1] 「河童」の「一」より抜粋

に主語が共通の場合は間を短くとる要因となる。また、後の文の主語を強調したい場合は、間を長くとる。例えば以下の2文である[2]。

1. 教場のしくじりが生徒にどんな影響を与えて、その影響が校長や教頭にどんな反応を呈するかまるで無頓着であった。
2. おれは前に云う通りあまり度胸の据った男ではないのだが、思い切りはすこぶるいい人間である。

2文のあいだの間は2450msecで、主語は「おれ」である(文1では省略されている)。間をあけることで、主語を「無頓着」であるが「すこぶるいい人間」であることを強調している。

4.2.3 段落

単語頻度ベクトルに対する内積は文書構造の把握に有益であることがHearstにより示されている[10]。そこで連続する2段落の関係を表すために、それぞれの単語頻度ベクトルを求め、その内積を求めた。

連続する2段落の関係を

- 2段落の内積が1に近い → 2段落の関係が近い
- 2段落の内積が0に近い → 2段落の関係が遠い

と定義する。そこで内積(横軸)と間の長さ(縦軸)の関係を調査した(図4)。事前に、2段落の関係が近ければ間が短くなり、遠ければ長くなるような、左上から右下へ減少の傾向がみられることを予想していた。「蜘蛛の糸」ではその傾向が観察される。しかし「坊ちゃん」「河童」ではそのような傾向を見ることができない。つまり、連続する2段落の関係と間の長さには相関関係がないことが観察される。

これには以下の理由が挙げられる。杉藤によると、話し手の間は聴き手が短期記憶を復習し、話を理解するために重要であることを述べている。そこで我々は人間の読み手が順次文書を朗読するとき、段落は構造が大きすぎて、聴き手の短期記憶に入りきらないと考えた。とすると、もし読み手が連続する段落によって間を変えることが可能であっても、聴き手はその細かい意味を理解することが不可能である。そこで段落レベルで間を制御する重要性はあまりないと考えられる。それよりも、読み手は文の間よりも長い間を段落で用いることにより、単に段落の最後を明示するに留まるのであろう。

5 おわりに

音声合成器の生成する音声をより表現豊かにするために、発話の間に注目して文書構造との関係の解析を行った。具体的には「文書内の2つの要素の関係が近いほど、そのあいだにはより短い間をとる」という仮説を立て、日本語の物語の朗読データを用いて調査を行った。まず最初に、文節、文、段落というように要素の大きさが大きいほど、後に続く間は長くなる傾向にある事が観察された。また、要素と要素の関係を定義し、間の長さへの影響を調査した。その結果、要素の関係は間の長さを変化させる要因であることが確かめられた。しかし、これは文節、文の場合だけであり、段落ではそういう傾向は観察されなかった。

今後の課題としては、朗読データ以外の発話データにも、このような関係が存在するのか調査する必要がある。また文書から適切な間の場所と長さを計算するアルゴリズムを提案し、音声合成器が生成する音声をより自然なものに近づけていくプリプロセッサの開発を目標としている。

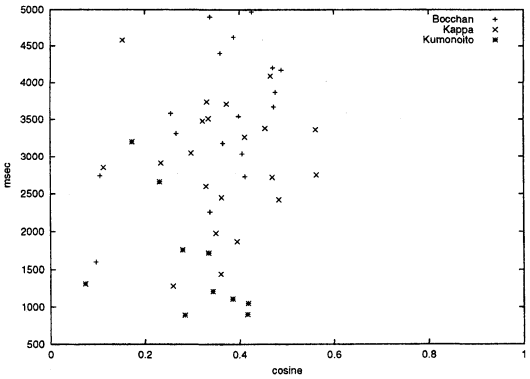

図4: 段落の関係と間の長さ

参考文献

[1] C. Nass and B. Reeves. Media Equation. Cambridge University Press, 1999.

[2] S.Pan and J.Hirschberg. Modeling Local Context for Pitch Accent Prediction. ACL, 2000.

[3] J.Fry. F_0 Correlates of Topic and Subject in Spontaneous Japanese Speech. ICASSP, 2000.

[4] 宮崎 正弘, 大山 芳史. 日本文音声出力のための言語処理方式. 情報処理学会論文誌, Vol.27 No.11, 1986

[5] 杉藤 美代子. 声に出して読もう. 明治書院,1996.

[6] 日本語読み上げソフト おしゃべりメイト. 富士通. http://www.fmw.co.jp/s11/oshaberi/

[7] Microsoft Agent Home http://msdn.microsoft.com/workshop/imedia/agent/default.asp

[8] 新潮CD「河童」「坊ちゃん」「杜子春」. 新潮社, 1997

[9] 黒橋 禎夫. 日本語構文解析システム KNP. http://www-lab25.kuee.kyoto-u.ac.jp/nl-resource/knp.html

[10] M.A. Hearst. Multi-paragraph segmentation of expository text. ACT, 1994

[2] 「坊ちゃん」の「三」より抜粋