E-051

音声合成による朗読システムに関する研究
Story Reading System by Speech Synthesis

吉田 有里† 奥平 康弘‡ 田村 直良†
Yuri Yoshida Yasuhiro Okudaira Naoyoshi Tamura

1. はじめに

近年、計算機技術の著しい発展にともない音声合成技術が進歩し、生成された合成音声が様々な場面で使用されるようになり、我々の生活に身近な存在となった。音声合成技術の応用例のひとつに、視覚障がい者や高齢者のための代読支援がある。実際、神奈川県内で一か月に300 件のリーディングサービスの利用があり、視覚障がい者からの利用が増加している[6]。しかし、現状では機械による音声を介して朗読を図るとき、人間が朗読するような自然で聞きとりやすいような発話は難しい。また、機械による無機質でポーズのない定常的な音声は、聞き手に対し不快感を与えてしまう要因となることもある。本来、ポーズの役割はテキストの内容を感覚的、意味的に捉えやすくすることであり、特に物語朗読においては、状況描写や臨場感、情緒的なつながりをもたせる。また、韻律は、変化に富んだ声調につながる。特に句末の韻律においては、疑問や強調など文のモダリティが話者の意図や態度を伝えるなど、重要な役割を果たしている。従ってこれらの要素は、テキスト音声合成による朗読には必須の技術である。

規則による音声合成を目指したポーズ挿入の分析・規則化は、かなり早い時期に行われている。比企[7]は呼気の影響によりポーズの継続時間（ポーズ長）が呼気にはさまれた発話区間の継続時間（呼気段落長）に比例することを示した。更に藤崎ら[8,9]は呼気との関係を考慮して、ポーズ長を呼気の伴わないポーズ、呼気を伴う文中および文末のポーズの 3 値（80ms、300ms、700ms）を用いることを提案している。これらの先行研究以降、音声器官の働きに基づく、呼気および呼気段落長が日本語におけるポーズ制御の重要な要因として考えられ、規則合成のポーズ制御にも用いられてきた。しかし、杉藤ら[10]は生理学的実験により、短い発話の後や短いポーズにおいて呼気を伴うポーズを挿入する場合があること、意味上の区切りとして重要な句境界では呼気を伴ったポーズが挿入されることを示した。このように呼気および呼気段落長にかかわらず、文の統語構造によりポーズ挿入特性がきまる場合があることが報告されている。また、海木ら[12]はポーズ挿入傾向・ポーズ長分布は話者により相違はあるが、性質の異なる長短の長短 2 種類のポーズが一般的に存在することを定量的に明らかにし、特定の句構造においてポーズが挿入され易いという知見が得られている。

中村[13]は心理学的観点から、ポーズと感情表現の間の明らかな関係を見出した。すなわち、ポーズの長さは、感情を込めて朗読した時の方が、感情を込めず淡々と朗読した時に比べて長いことである。

これらの従来研究では、句境界における係り受け関係を主要因としてポーズ規則が提案され、文間に存在するポーズの挿入規則化の定量的で詳細な分析はあまり行われていない、文間ポーズと当該文の韻律との関係が明確化されていないなどの問題点がある。

また、韻律に関して、文末の音調の分類を提案した研究が行われている。近年、服部[11]は、終助詞によるイントネーションの変化の種類を提案し、石井ら[14]は、日常会話における句末の種類による、句末音調がもたらす役割を分析した。しかし、これらの従来研究では、物語の朗読音声における韻律的特徴の規則化の検討が不十分である。本来これらは文脈情報と加味して制御されるべきである。

本研究では、朗読音声における文間ポーズ、韻律の特徴量の適切な制御規則の構築を目指し、より「自然で聞きやすい音声合成」手法を提案する。また、システムにより生成された文間ポーズ及び韻律を付与した音声について、被験者が感じた自然性を回答する聴取実験により、本手法の有効性を検証する。

なお、本研究では、小説や随筆、詩など朗読文全般を対象とし、パラメータの採取には「羅生門」、「走れメロス」、「伊豆の踊子」、「駈込み訴え」、「芋粥」のテキスト[17]及び音声データ[1,2,3,4,5]を用いる。

2. 機械学習を用いたポーズ判定

2.1 朗読文の文間ポーズ長の頻度分布調査実験

朗読者が発声した音声を音響分析し、朗読文の文間ポーズ長を抽出し、その頻度分布を見た。

その結果から、朗読音声には長短のポーズの分布が見られることがわかった。

つまり、短ポーズの分布、長ポーズの分布が重なって一つの頻度分布を構成しているように見てとれる。

本研究では、これら短ポーズの分布、長ポーズの分布がそれぞれ正規分布であると仮定し、それぞれの長短ポーズの分布を推定する。

具体的には、短ポーズの分布を短ポーズの正規分布、長ポーズの分布を長ポーズの正規分布にあてはめ、長短ポーズの正規分布から成る二重混合正規分布による近似を行う。

二重混合正規分布の推定に必要なパラメータは、平均、分散、混合比であり、これらのパラメータを EM アルゴリズム[19]を適用することにより推定する。

羅生門についての文間ポーズ長の頻度分布と、それから算出された二重混合正規分布を図 1 に示す。

2.2 判定器の構築

素性抽出に用いるデータとして「羅生門」、「走れメロス」、「伊豆の踊子」、「駈込み訴え」、「芋粥」の

† 横浜国立大学環境情報学府
‡ 出光興産株式会社

全文 2145 文を使用する。本研究では、15 個のパラメータを用いた。（表 1）

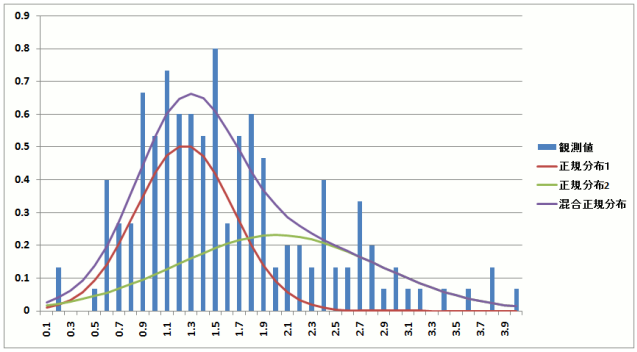

図 1　羅生門における文間ポーズ長の混合正規分布

表 1　素性抽出パラメータ

カテゴリ	パラメータ	値
句点の前文	モーラ数	数値
	文形式	1/-1
	指示詞の有無	1/-1
	主題の有無	1/-1
	場所表現の有無	1/-1
	時間表現の有無	1/-1
句点の後文	モーラ数	数値
	文形式	1/-1
	指示詞の有無	1/-1
	主題の有無	1/-1
	場所表現の有無	1/-1
	時間表現の有無	1/-1
句点を挟む前文・後文	主題変化の有無	1/-1
	場所転換の有無	1/-1
	名刺の語彙連鎖値	数値

2.3 判別実験

二重混合正規分布において、2 つの正規分布の交点の時間を 4 作品について平均し、この値より長いポーズを長ポーズ、短いポーズを短ポーズとして、C5.0 により機械学習した。

判別結果を以下の表 2 に示す。

表 2 クローズテスト及びオープンテストの結果

Close test		Predicted	
		短い	長い
Actual	短い	92%	8%
	長い	68%	32%
Accuracy		76.9%	

Open test		Predicted	
		短い	長い
Actual	短い	89%	11%
	長い	68%	32%
Accuracy		75.1%	

3. 韻律の傾向とモデル化

3.1 朗読文の韻律調査実験

本研究では、実際に朗読者が発声した音声を音響分析し、基本周波数・話速の韻律変化パターンを朗読対象文に適用する。そのため、本節では朗読者の音声の韻律調査を行った。

実験に使用した機器

音声データを解析するためのフリーソフトウェアである wavesurfer-185-win[21]を使用する。

実験対象

朗読文の音声データ 1123 文を用いる。その内訳は以下である。

- 「羅生門」130 文
- 「走れメロス」292 文
- 「駈込み訴え」393 文
- 「芋粥」308 文

結果

分析により、韻律の変化が表れている箇所について、基本周波数については一文の最終文節の最終音素に基本周波数の上げ下げがあり、話速については一文全体に変化があった。なお、最終音素のイントネーションの型（音調）については、上昇調、平調、下降調の 3 つに分けた。

基本周波数についての結果を表 3 に、話速についての結果を表 4 に示す。

表 3　基本周波数の変化分析結果

	上昇調	平調	下降調
羅生門	46文	32文	52文
走れメロス	99文	53文	140文
駆け込み訴え	133文	76文	184文
芋粥	68文	67文	173文

表 4　話速の変化分析結果

	速い	普通	遅い
羅生門	27文	55文	48文
走れメロス	163文	113文	16文
駆け込み訴え	188文	160文	145文
芋粥	48文	148文	39文

3.2 韻律変化パターンへの検討

以上の調査より、朗読文の特徴として、動作や程度といった文内表層情報や文末表層情報によって変化が見られる場合が多いことが分かった。また、文内表層情報と文末表層情報による韻律の変化を比較した場合、文内表層情報の方が影響が強い場合が多いことが分かった。さらに文内表層情報においても、動作や程度などのカテゴリ別において、その韻律への影響の強さに違いが見られた。そこで本研究では、文内表層情報、文末表層情報に基づく韻律のモデル化を行うことにする。

3.3 文内、文末表層情報の一般化

3.3.1 文内表層情報の一般化

文内表層情報から文を命令、否定、意志、推定、理由、副詞、程度、怒、悲、動作、悪、感動詞、接続の 13 のカテゴリに分類する。文内表層情報は重複する場合がある。

本研究では朗読音声の韻律結果から、次の順序関係「命令>否定>意志>…>悪>感動詞>接続」を設定する。

3.3.2 文末表層情報の一般化

朗読音声中に出現した文末形式の内、多く現れた形式を文末表層情報と定義し、「～ある」、「～いる」、「～んだ」、「～だ」、「～だろう」、「～あろう」、「～らしい」、「～あった」、「～いた」、「～のだ」、「～やろう」、「～やる」、「～やれ」、「～れた」、「～ない・～ぬ」、「～たい」、「～なった」、「～だからな」、「～くれ」、「以外」の 20 のカテゴリに分類する。

3.4 韻律付与システム

解析により文内表層情報、文末表層情報を得る。もし、文内表層情報の表現が有れば韻律パターンを適用し文全体に話速を、最終音素に基本周波数を付与する。なければ文末表層情報の表現により韻律パターンを適用する。

文に付与する話速のパターンは遅くするか、速くするかの 2 通りで、最終音素に付与する基本周波数のパターンは上げるか下げるかの 2 通りである。

4. 音声合成システムの評価と考察

構築した朗読システム(図 4)[1]から文間ポーズと韻律（基本周波数、話速）を付与された朗読システムの合成音声と音声合成器のオリジナル音声を比較し、朗読の雰囲気を知覚するか、また、自然性が失われていないかについて聴取実験を行い、本手法の評価を行う。

実験には本朗読システムの合成音声と Visual Speech Creator[18]によるオリジナル音声を用い、知覚実験による評価を行う。

4.1 実験１：朗読の雰囲気や状況に適した朗読

朗読システムによる合成された朗読システムの合成音声が、朗読の雰囲気や状況にあった朗読になっているかを調べる。

実験に用いる朗読音声データは 83 文である。朗読音声は、ほとんどの朗読文に存在するような場面を考慮した「状況描写」（羅生門）に関する 9 文、「緊迫」（羅生門）に関する 10 文、「天候」（羅生門）9 文、「人物紹介」（走れメロス）10 文、「災害」（走れメロス）9 文、「自己奮起」（走れメロス）13 文、「告白」（駈込み訴え）14 文、「不安」（芋粥）9 文を選択し、それぞれ 8 つにカテゴリ分けする。被験者には、各カテゴリに場合分けされた場面について回答させる。実験には、それぞれの場面をあらかじめ提示し、ランダムに順序付けた朗読システムの合成音声とオリジナル音声を 1 組とし計 8 セットについて、「より場面に適切な音声をどちらか」選択させる。被験者は日本語話者 3 名である（以下同様）。

4.2 実験２：合成音声の自然性

朗読システムの合成音声の自然性が損なわれていないかを調べるため、朗読システムの合成音声における文間ポーズ、韻律の適切性を調べる。

実験に用いる朗読音声データは、83 文である。実験には、ランダムに順序付けた朗読システムの合成音声とオリジナル音声を 1 組とし、計 8 セットについて、自然で聞き取りやすい音声をどちらか選択させる。

4.3 実験３：文単位における文間ポーズ、韻律の評価

文単位における文間ポーズ、韻律の変化の印象度を調べる。実験に用いる朗読音声データは、8 文章（83 文）である。実験は、各文章の文単位の朗読システムの合成音声について、適切であるか、不適切であるかの印象度を文間ポーズ、韻律について選択させる。また、不適切な場合には、文間ポーズについては、「長く感じる」か「短く感じる」かについての回答を、基本周波数については「上ずっているように感じる」か「下っているように感じる」かについての回答を、話速については「速く感じる」か「遅く感じる」かについてそれぞれ回答させる。

4.4 結果と考察

- 実験 1・実験 2

  実験１、実験２の結果を表 5 に示す。

表 5　実験1, 2結果

	朗読システムの合成音声	オリジナル音声
実験1	87.5%	12.5%
実験2	79%	21%

オリジナル音声よりも朗読システムの合成音声の方が適切だとする割合が高いことにより、朗読の雰囲気や状況に適した朗読であることや、より自然な朗読音声であることが示された。また、実験 1 でオリジナル音声の方が好ましいとしたカテゴリについては、「緊迫」、「自己奮起」、「不安」が見られた。これは「緊迫」や「自己奮起」、「不安」といった雰囲気や状況が徐々に声調が変化して朗読されるのが一般的であり、例えば、「緊迫」の場面では、徐々に緊張や、圧迫感が差し迫る音声を再現する必要があると考えられる。

- 実験 3

  実験３の結果を表 6、7 に示す。

表 6　実験3の結果

		文間ポーズ	基本周波数	話速
実験3	適切率	79.1%	78.3%	92.7%
	不適切率	20.9%	21.7%	7.3%

表 7　場面ごとに集計した実験3の結果

	状況描写	緊迫	天候	人物紹介	災害	自己奮起	告白	不安
文間ポーズ	75%	85.1%	62.5%	88.8%	83.3%	80.5%	74.3%	83.3%
基本周波数	59.2%	80%	70.3%	93.3%	77.7%	72.2%	90.4%	85.1%
話速	100%	93%	96.2%	100%	81.4%	92.3%	88%	92.5%

まず文間ポーズであるが、「状況描写」や「天候」、「不安」において評価が悪かった。その原因の多くがポーズが長いというものであった。その主な理由として「状況描写」に関しては意味的に切れる部分でもポーズが長かった箇所が多かった。また、「不安」に関しては『まして』といった前文を補足するような内容が続くこ

[1] ポーズ長のばらつきの再現が困難であったため、長ポーズの分布、短ポーズの分布を正規分布乱数を用いることで再現している

とを示し、文間のつながりを強くする接続詞があるのに、長ポーズを設けてしまっていた。この事から場所などの焦点があたっている対象に関する文が続く時は、ポーズを短く、焦点が当たる対象が切り替わる時はポーズを長くするなどの意味を考慮した調整が必要であると考えられる。次に基本周波数であるが、「状況描写」や「天候」、「災害」で評価が悪かった。その原因の多くが、全体的に上ずった音声になっているというものであった。これは、元々上がっていた音声をさらに上げてしまった事、また、最終音素のみを調整したことでその直前の文との基本周波数とのつながりが不自然になってしまったことなどが原因だと考えられる。

全体の結果に対する考察として、被験者は 80%前後、オリジナル音声よりは自然であると回答し、ポーズ、韻律の付与に関しても適切であると回答している。特に話速に関しては、90%程度適切であると回答しており、本手法による有効性が確認された。しかし、ポーズ、基本周波数については、より有効なパラメータを考慮し制御する必要があるといえる。

5. おわりに

本研究では、朗読文を対象に音声合成器の合成音声の自然性を高めることを目的とし、朗読者の音声の間（ポーズ）及び、韻律的特徴（基本周波数、話速）を解析し、発話の特徴のモデル化を行った。また得られたモデルを用いて、ポーズ及び韻律の変化を実現する朗読システムを構築した。合成した朗読音声の聴取実験の結果、朗読文にポーズや韻律の変化を付与することには一定の効果があったと言える。

今後の課題としては、文間ポーズの判別の向上がある。意味を考慮したパラメータを用いるなど、より大局的な構造を検討する必要がある。

また、韻律に対して用いた適用範囲を拡張する必要がある。最終音素に限らず、全ての音素で複雑な韻律の変化が見られるため、まずは文節レベルで拡張し、韻律変化を検討する必要がある。

最後に、朗読文章から喜怒哀楽といった感情を読み取り、付与することで、朗読の臨場感や雰囲気を向上させることができるのではないかと考えられる。

参考文献

[1]朗読・橋詰功・芋粥,新潮社,2002
[2]朗読・橋詰功・羅生門,新潮社,2002
[3]朗読・篠田三郎・伊豆の踊子,新潮社,2002
[4]朗読・草野大悟・駆け込み訴え,新潮社,1997
[5]朗読・草野大悟・走れメロス,新潮社,1997
[6]間野和則,金子孝夫,高品質音声圧縮技術を用いた視覚障害者のための朗読配信システム,電子情報通信学会論文誌,Vol..J83-D-1,No.11(2000)
[7]比企静雄,連続音声中の各種区分の持続時間の性質,信学誌,Vol.50,No.5(1967)
[8]藤崎博也,韻律研究の諸側面とその課題(1994)
[9]藤崎博也,広瀬啓吉,合成音声とアクセント・イントネーション,信学誌,Vol.70,No.4(1987)
[10] 杉藤美代子,大山玄,朗読におけるポーズと呼気—息継ぎのあるポーズと息継ぎのないポーズ,(1990)
[11] 服部匡,終助詞の音調について,第 14 号(2002)
[12] 匂坂芳典,海木延佳,局所的な句構造によるポーズ挿入規則化の検討,電子情報通信学会論文誌,No.9(1996)
[13] 中村敏江,「間」の感性情報,日本ファジイ学会誌,vol.14,No.1
[14] 板橋秀一編著,赤羽誠,石川泰,大河内正明,粕谷英樹,桑原尚夫,田中和代,新田恒雄,矢頭隆,渡辺隆夫共著,音声工学,森北出版(2005)
[15] 形態素解析システム・茶筌, http://chasen.naist.jp/hiki/ChaSen/
[16] 係り受け解析システム CaBoCha, http://chasen.org/~taku/software/cabocha/
[17] 青空文庫,http://www.aozora.gr.jp/
[18] Visual Speech Creator, http://www.ntt.it.co.jp/goods/vcj/voice/vsc.html
[19] http://www.ntt.dis.titech.ac.jp/sekino/paper/note/EMalgorithm.pdf
[20] Fine Voice, http://www.ntt.it.co.jp/goods/vsj/voice/finevoice.html
[21] Wavesurfer, http://www.speech.kth.se/wavesurfer/

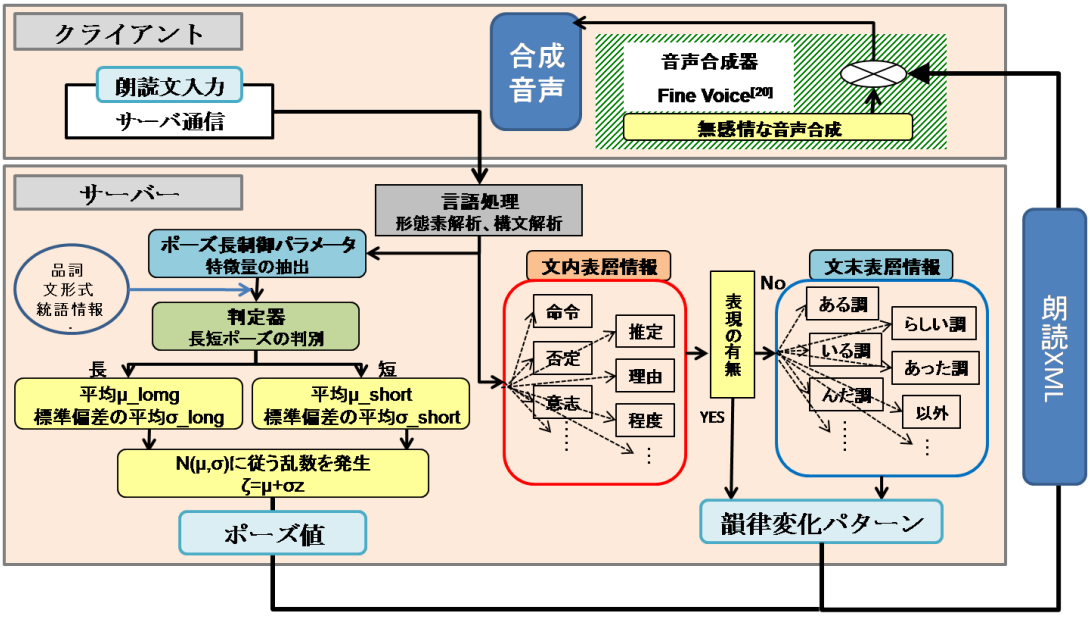

図2　自動朗読システム