テキスト音声合成における係り受け解析結果を用いたポーズ挿入処理

佐藤 奈穂子[†] 小島 裕一[†]
望主 雅子[†] 亀田 雅之[†]

我々は、テキスト音声合成システムのポーズ挿入精度向上のために、言語処理部に構文解析処理を導入し、一文全体の係り受け解析結果を利用したポーズ挿入処理を試みた。本稿では、この結果について報告する。テキストを音声に変換して出力する際には、その内容を感覚的、意味的に捉え易くするために、テキスト中の適当な位置に適当な長さのポーズを与える必要がある。ポーズ位置やポーズ長の設定に、対象文の係り受け解析結果が有効な手がかりになるとの知見が得られているが、従来は実施上の都合から、語彙情報の利用や、局所的なテキスト解析による方法が代用されていた。そこで本稿では言語処理部に軽量かつ高速な構文解析系を導入し、一文全体の係り受け解析のシステム上での実現を試みた。ポーズ挿入の生じ易さの指標としてポーズ挿入尤度を設け、係り受け情報に着目したポーズ挿入規則に基づき、全文節境界にポーズ挿入尤度を設定する。尤度の高い境界から基本ポーズ長レベルを設定した後、各境界に対してアクセント結合処理および呼気段落に基づく閾値による調整を行なう。実際にテキスト音声合成システムに実装し、形態素解析と隣接間係り受け処理のみ実装しているテキスト音声合成ソフトウェアパッケージとの比較実験を行なったところ、ポーズ挿入精度の大幅な向上が得られ、その効果を確認することができた。

キーワード: 音声合成, ポーズ挿入, 係り受け解析

Pause Insertion which uses the results of Grammatical Dependency Analysis within a Text-to-Speech Conversion System

NAHOKO SATO[†], YUICHI KOJIMA[†], MASAKO MOTINUSHI[†]
and MASAYUKI KAMEDA[†]

This is a practice of method which uses results of grammatical dependency analysis within a text-to-speech conversion system to insert pauses. In order to understand the meaning of text which is converted into speech, the appropriate insertion of pauses into phrase boundaries should be required. In previous studies, several approaches using the results of simple analyses such as morphological analysis or analysis of adjoining phrases to determine the position and length of pauses was proposed. In those studies, we often found the speech had inappropriately determined pause positions. In the present method, we introduce a quick Japanese parser into the text analysis, determine the distance and relationship between dependent phrases, and use this information to determine the position and length of pauses. The distance and relationship between dependent phrases is translated into the cost of pause insertion. Each phrase boundary has a cost of pause insertion, and appropriate pauses are inserted according to the costs. To test the validity of our method, we implemented

it in a text-to-speech conversion system, and compared the proposed method with a previous pause insertion method which used simple analyses such as morphological analysis or analysis of adjoining phrases. The results confirmed that our proposed method fulfilled our expectations.

KeyWords: *Speech synthesis, Pause insertion, Dependency analysis*

1 はじめに

テキスト音声合成システムの言語処理部における重要な課題の一つに、ポーズ挿入処理が挙げられる。ポーズ挿入処理は、音声化され、出力されたテキストの内容を人間が感覚的、意味的に捉えやすくするために、テキスト中の適当な位置に適当な長さのポーズを与える、テキスト音声合成に必須の技術であり、入力テキストの書き手が意識して挿入した句読点以外にも構文構造とポーズ挿入位置の関係が研究されてきた。

従来の研究から、ポーズは構文的区切りと一致する(杉藤 1988)、また特定の句構造においてポーズが挿入され易い(海木, 匂坂 1996)という知見が得られている。この他にも、文節間の係り受け距離と文節の長さが、ポーズ挿入の有効な手がかりになるという知見(箱田, 佐藤 1980)、さらに、係り受け関係、句読点、文中における位置情報を加えることで精度が高まると期待できるとの報告(箱田他 1989)もある。しかし、これらは係り受け距離や係り受け関係などのテキスト情報が既に得られているとの前提に立った報告であり、実際にそれらのテキスト情報を求めるためには構文解析処理が別途必要となる。

一般に構文解析処理は、大量の言語知識データを要する、テキストから精度の高い統語構造の自動抽出が困難、処理が重くなる、などといった問題から、実働するシステムにおいては、簡易なテキスト解析で得られる単語の品詞やモーラ数など、形態素解析レベルで得られる情報や、局所的な数文節に着目した簡易な係り受け解析が広く用いられている(宮崎, 大山 1986; 浅野他 1995; 鈴木, 斎藤 1995; 澗瀧他 1996; 塚田, 田中 1996; Tsukada 1996; 海木, 匂坂 1996, など)。係り受けの範囲については、隣接する数文節の範囲内に限定できるとの報告(箱田他 1989; 鈴木, 斎藤 1995)、実際の文章において、隣り合う2文節の係り受けが連続するケースが多いという報告(丸山, 荻野 1992; 張, 尾関 1997)があり、隣接2文節、もしくは局所的な数文節間の係り受け解析結果を用いた方法でかなり高精度のポーズ挿入が実現できることが明らかになっている。

しかしながら、人間が聞いて理解しやすい構文的まとまりは、複数の文節によって様々なパターンで構成されており、上記方法でも限界はある。例えば、小説や随筆など、一文がある程度長く、文の構造が複雑なものになると、係り受けが3文節以上に跨る文の存在は少なくない。予め係り受けの範囲を3文節に限定してしまうことで、構文的まとまりの一節中にポーズが挿入されるなど不自然な読み上げを頻出する場合がある。

一方、別のアプローチの一つに、コーパスを利用した統計的なポーズ挿入位置の予測方法が

† (株)リコー 情報通信研究所, Information and Communication R&D Center, RICOH Co., Ltd.

報告されている。文献(Iwata et al. 1990)では、隣接2単語の接続のしやすさをコーパスを用いてスコア化し、それを用いたポーズ挿入方法を提案している。また、文献(Doi et al. 1994)では、副助詞や接続助詞などの文法的役割に着目し、コーパスを用いてそれらの語彙の後に来るポーズの長さをレベル化し、それを用いたポーズ挿入方法を提案している。さらに文献(藤尾他1997)では係り受け情報付きコーパスの学習とポーズ情報付きコーパスの再学習によりフレーズ境界前後の形態情報とポーズ長の関係を統計的に得る方法を提案している。しかし、これらの方法は予め大量の学習用データを要し、さらにデータの分野依存が大きいと考えられる。

本稿では、大量の学習用データに頼らず、長距離の係り受け解析をする、軽量・高速な構文解析処理を用いたポーズ挿入手法について報告する。本手法では、解析の範囲を文の長さや文節数で限定せず、一文を単位とした係り受け解析の情報を利用する。また、本手法をPC上で実動するレベルのテキスト音声合成システムに実装して、その効果を確認した。

2 構文解析処理の導入

2.1 利用する構文解析処理系

本手法では、日本語解析系Quick Japanese Parser (以下、QJP) (亀田 1996; Kameda 1996, など)をテキスト音声合成システムの構文解析に利用する。先に述べたように、一般的な構文解析で、実用レベルの開発を難しくしている要因に、詳細かつ膨大な(場合によっては意味まで踏み込んだ)規則を要する、生成される候補が多い、計算量も多い、などの問題がある。また、その解析精度も形態素解析技術に比べると処理が複雑であるため、なかなか実用レベルになりにくい。しかし、QJPは、日本語解析を要する応用系システムへの組み込みが容易な言語処理ライブラリであり、効率的な処理が実現可能である。

QJPは字種の特徴を利用した小規模辞書ベースの形態素解析系と、形態素情報ベースの規則による構文解析系の2つの処理系で構成されている。主に形態素情報による規則を用いて解析処理を行なうため、シソーラスや意味情報など、構築およびメンテナンスに労力を要する膨大な言語知識データは不要であることを大きな特徴としている。構文解析系は、一文毎に単語列を文節にまとめる際、文節にはその構成単語の品詞情報を利用して文節属性が設定される。その文節属性や品詞情報に基づき係り受け可能文節対を検出し、非交差規則、再近接文節選択をベースに、様々なヒューリスティックを用いた例外処理も用いて、尤もらしい係り受け文節対を確定する。品詞情報とその組み合わせで構文解析処理を行なうため、解析規則等のデータ開発コストが抑えられ、軽量さ、高速さを実現できている。また、長文に対しての頑健さも兼ね備えているため音声合成向けのテキスト解析に導入しやすい。さらに、係り受け処理の単位が一文であり、文節数が限定されないため、対象となった文中の本来の文法的区切り位置を得ることが可能である。

2.2 ポーズ挿入のための利用可能情報

構文解析処理を導入することで、利用可能なテキスト情報が大幅に増える。これらの情報をどのように利用するかで様々な手法が考えられ得る。

まず、形態素解析結果から得られるポーズ挿入処理に有益なテキスト情報は、以下の通りである。

- 句読点位置情報
- 語彙レベルの定性的情報

感動詞、接続詞、提題の意味の「は」の直後にはポーズが入りやすい、などの品詞や語特有の性質

- 文節モーラ長
- 局所的な連続文節間の関係

隣接を含む局所的な連続した数文節間における係り受け関係の有無

さらに、QJPの構文解析系の解析結果から得られるポーズ挿入処理に有益なテキスト情報は以下の通りである。

- 文節の属性
- 係り受け文節対
- 係り受け文節間の距離

隣接文節間の係り受け距離を1としてカウントする（一文中の文節数 > 距離 > 0）

- 係り受けの関係
係り受け文節対の各属性に基づく

そこで、まず予備調査として、先行研究(箱田, 佐藤 1980; 箱田他 1989, など)で取り扱われている、係り受け文節間の距離を用いてポーズ挿入実験を行ない、その評価結果に基づいて実際のポーズ挿入処理アルゴリズムの設計を行なうことにした。

2.3 予備実験

本手法の予備調査のために、係り受け文節間の距離のみを用いたポーズ挿入実験を行なった(佐藤 1997)。以下、予備実験について述べる。

2.3.1 実験方法

実験は下記の手順で行なった。使用したサンプルテキストは日本電子工業振興協会の報告書(日本電子工業振興協会 1995)の付録の評価文の抜粋および小説(村上 1995)の抜粋を用いた。また、比較対象は形態素解析と隣接間係り受け処理を用いたポーズ挿入処理を実装している既

製のテキスト音声合成ソフトウェアパッケージ¹の出力である。

- (1) サンプルテキストを比較対象のテキスト音声合成ソフトに入力し，ポーズ記号，アクセント句区切り記号を含む発音記号列を自動的に作成する
- (2) 同ソフトにおいて同じサンプルテキストを発音記号列に自動的に変換した後，挿入されたポーズ記号，アクセント句区切り記号を取り除き，「読み」のみの記号列を作成する
- (3) Q J Pに同じサンプルテキストを入力し，形態素解析&構文解析結果を得る
- (4) Q J Pの出力結果から係り受け文節を抽出し，係り文節から受け文節までの距離を表1の通り各ポーズレベルへ対応づける
- (5) 予め作成しておいた(2)の読み記号列に(4)の結果を反映させ（対応ポーズレベルをポーズ記号に変換して手入力），文節間距離情報を反映した発音記号列を作成する
- (6) 比較対象のテキスト音声合成ソフトが自動作成した(1)の発音記号列と，文節間距離情報を反映した(5)の発音記号列を比較・評価する

表 1 文節間距離-対応ポーズレベル

距離	ポーズレベル
距離 1（直後に係る）	アクセント句区切り
距離 2	小ポーズ
距離 3	中ポーズ
距離 4 以上	大ポーズ

2.3.2 結果

評価はポーズ挿入正解率，ポーズ挿入精度向上率，ポーズ挿入精度降下率という3つの観点で行なった。

ここで言うポーズ挿入正解率とは，期待ポーズ位置数に対して期待通りのポーズが挿入された割合，ポーズ挿入精度向上率とは，比較対象のテキスト音声合成ソフトで失敗した箇所が文節間距離情報を反映した方法で成功している割合，ポーズ挿入精度降下率とは，比較対象のテキスト音声合成ソフトで成功していた箇所が文節間距離情報を反映した方法で失敗した割合と定義している。評価結果を表2に示す。

1 (株)リコー 製テキスト音声合成ソフトウェア「雄弁家V2」を使用

表 2 評価結果

ポーズ挿入正解率 (比較対象のテキスト音声合成ソフト)	92.0 %
ポーズ挿入正解率 (文節間距離情報を反映した方法)	97.6 %
ポーズ挿入精度向上率	94.8 %
ポーズ挿入精度降下率	2.2 %

2.3.3 予備実験における考察

評価結果より、係り受け文節間の距離だけでも形態素解析と隣接間係り受け処理を用いたポーズ挿入処理を実装したシステムのポーズ挿入誤りを大幅にカバーできるということが明らかになった。

一方、ポーズ挿入精度降下の原因を調査すると、多くはQ J Pの形態素解析誤りのため、構文解析処理に失敗し、正確な係り受け文節間の距離が得られなかったことに起因していた。しかし、正確な係り受け文節間の距離が得られても、以下のような問題点が明らかになった。

●入力テキスト

「複雑な技術だけが生み出すことのできる種類の優美さだった。」

●実験結果出力された発音記号列²

フクザツナ/ギ' ジュツダケガ/ウミダ' ス/コト' ノ/デキ' ル/
シュ' ルイノ/ユービ' サダッタ.

上記例文では、各文節が直後に係る (距離 = 1) 係り受け関係が連続しており、各アクセント句境界に同じアクセント句区切りが挿入されてしまった。このような場合、読み上げが単調になるため、聞き手側には構文的区切り位置がどこにあるのか分からないなど、不自然さが残る。

構文解析により構文的区切り位置の情報を得ることができても、実際には上記例文のように、距離 = 1 の係り受けが連続する場合が多い (丸山, 荻野 1992; 張, 尾関 1997) ため、係り受け文節間の距離を単純にポーズ長へマッピングするだけでは、実装上まだ不足であり、ポーズ長の制御のための何らかのテキスト情報を取り入れる必要がある。

そこで、次に、文献 (箱田他 1989) を受けて、テキスト情報として係り受け文節間の距離に加えて、係り受けの関係を利用することにした。さらに、語句や呼吸段落における定性的な規則も用いることにした。

² 発音記号列における「/」はアクセント句境界またはポーズ挿入位置を示す。また、ここでは「/」の数が1のときはアクセント句境界を表し、2以上のときはポーズを表す。2以上の場合、数が多いほどポーズの長さが長いことを示す。

2.4 ポーズ挿入規則

本稿で提案するポーズ挿入手法では、基本的にテキスト中の各文節末には何らかの韻律句境界があると定義する。そして、各韻律句境界にポーズ挿入規則に基づき、ポーズ挿入が生じ易いかどうかの指標をポーズ挿入尤度として設定する。規則により全ての韻律句境界にポーズ挿入尤度が設定されたら、ポーズ挿入尤度の高い境界から順に、適当なレベルのポーズを挿入する。

ポーズ挿入規則は、文献(箱田他 1989)に基づき、係り受け文節間の距離に基づく規則と係り受けの関係に基づく規則と、語句や呼気段落における定性的な規則の3規則にまとめた。

(1) 係り受け文節間距離(d)に基づく規則

係り受け文節間距離(d)とは、着目している文節位置直後から、その係り先文節までに含まれる文節の数(係り先文節を含む)とする。従って、隣接文節は $d=1$ となる。dはそのままポーズ挿入尤度とする(表3)。テキスト中にポーズを挿入する際は、尤度に対応したポーズ長レベルを予め設定しておく必要がある。ポーズ長レベルはシステムの用途によって変更可能である。

表 3 規則 1

文節間距離(d) = ポーズ挿入尤度	ポーズ長レベル設定の一例
$d = 1$	アクセント句区切りレベル
$d = 2$	声だて(立て直し)レベル
$d = 3$	小ポーズレベル
$d = 4$	中ポーズレベル
$d > 4$	大ポーズレベル

(2) 係り受けの関係による規則

予備調査の結果、上記規則1だけでは、ポーズ長が一定で単調、構文のまとまりがわかりにくい読み上げとなる場合があることが明らかになった。そこで、本手法では規則1で $d=1$ の場合に限り、更に係り受けの関係に基づくポーズ挿入尤度を設定することにした。QJP構文解析系の係り受けの関係はかなり多くのバリエーションがあるが、本手法ではこれらを文法に即して大まかに6種に大分類し、対応するポーズ挿入尤度を持たせた(表4)。

表 4 規則 2

係り受け関係(大分類)	ポーズ挿入尤度
係り先が文末文節	Q
隣接関係	N
複合関係(アクセント結合の可能性あり)	D
連体関係	T
並列関係	H
連用関係	Y

ポーズ挿入尤度の高低は隣接文節関係の知見より以下のように設定した。

D - T - H - Y - N - Q

尤度(低) → 尤度(高)

規則 2 ではポーズ挿入尤度はアルファベット 1 文字のフラグで表記しており、規則 1 同様、テキスト中にポーズを挿入する際は、尤度に対応した基準のポーズ長レベルを予め設定しておく必要がある。

(3) 語句や呼吸段落における定性的な規則

句点の位置には文末相当のポーズを挿入する、読点の位置にはそれ相応のポーズを挿入する、など従来研究の知見や形態情報から得られる定性的な規則も併用する。これらは文献(箱田他 1989)で挙げられているテキスト情報「句読点」「文節位置」の他、文節を構成する単語の語彙情報も含んでいる。

また、テキスト中にポーズ挿入尤度の高い境界から順に、実際に適当なレベルのポーズを挿入した際、同レベルのポーズに挟まれた発話区分のモーラ数が呼吸段落を考慮して予め任意に設定した一定のモーラ長を超過した場合は、その間にある各文節境界のポーズ挿入尤度を参照し、最もポーズ挿入尤度が高い境界位置のポーズ長レベルを格上げし、同レベルのポーズを挿入する。

2.5 ポーズ挿入処理のアルゴリズム

2.4節で述べたポーズ挿入規則に基づき、入力テキストに以下の手順でポーズ位置とポーズ長を設定する。

(1) 文節間距離の算出

係り受け成立が確定した文節間の d を算出する

(2) d に対応するポーズ挿入尤度設定

規則 1 に基づき、各文節末 (= アクセント句末) へ d に応じたポーズ挿入尤度を設定する

(3) 係り受けの関係によるポーズ挿入尤度設定

$d = 1$ の係り受け文節間にのみ、規則 2 に基づき、係り受けの関係によるポーズ挿入尤度 (基準値) を設定する

(4) 句読点位置へのポーズ挿入尤度設定

句点の位置に文末相当の尤度を、読点の位置に適当な尤度を挿入する

(5) 各境界へのポーズ長の設定

各文節末にポーズ挿入尤度が設定されたら、アクセント結合処理などを経て、ポーズ挿入尤度の高い境界から順に予め設定しておいた尤度に対応した適当なレベルの長さのポーズを挿入する

- (6) 発話区分のモーラ長限界によるポーズ長レベルの格上げ
同レベルのポーズに挟まれた発話区分のモーラ長が呼気段落を考慮して予め任意に設定した一定のモーラ長を超過した場合は、その間にある各文節境界のポーズ挿入尤度を参照し、最もポーズ挿入尤度が高い境界位置のポーズ長レベルを格上げして、同レベルのポーズを挿入する

3 システムへの実装

前節で述べたポーズ挿入処理を、実際にPC上で実動するテキスト音声合成システムへ実装した。

3.1 システム構成

QJPの形態素解析系では、単語辞書を利用せず、字種の特徴を利用して単語切りをしており、音声合成に必要な読み・アクセント情報を得る手段が無い。また、漢字複合語列を各単語に分割しないため、後処理にアクセント結合処理が控えている音声合成の言語処理では問題になる。そこで、本システムへの実装にあたり、形態素解析には読みやアクセント情報記載の単語辞書を利用する形態素解析モジュールを用い、その出力をQJPの構文解析モジュールへ入力できるようなシステム構成を考えることにした。

形態素解析処理に用いる品詞の体系と構文解析処理に用いる品詞の体系が異なる場合、形態素解析処理の出力結果をそのまま構文解析処理の入力とできないことがしばしばある。そのような場合には、間に品詞変換処理が必要となるが、本システムでも同様に、QJP構文解析系を利用するために、品詞変換モジュールを併せて開発した(望主, 亀田 1998)。本システムでは構成上、品詞変換モジュールとQJP構文解析モジュールを合わせて、拡張QJP構文解析系と呼ぶ。また、後続処理である、アクセント結合モジュールとポーズ挿入モジュールを合わせてフレージング処理系と呼ぶ。

本システム構成を図1に示す。

図 1 システム構成

4 発音記号生成実験

本手法の効果を確認するために、簡単な評価実験を行なった。
以下、実験内容について述べる。

4.1 サンプルテキストと比較対象

評価実験にはポーズの正確さを評価するため、一文がある程度長く、文の構造が複雑である小説(村上 1995)から引用した文を入力用サンプルテキストとして用いた³。一文の平均文節数は21.5、平均モーラ数は146.5である。

3 予備調査の実験サンプルと出典は同じであるが、引用文に重複はない

比較対象は 2.3.1 節の予備調査の実験と同様のテキスト音声合成ソフトウェアパッケージ（以下、従来システムと呼ぶ）の出力した発音記号列である。

4.2 実験方法

実験は以下の手順で行なった。

- (1) 形態素解析誤りが構文解析処理に影響しないように、必要に応じて従来システム、および本手法を実装したシステムへの単語登録、解析誤り訂正を予め行なっておく
- (2) サンプルテキストに対し、最も自然に聞こえるように人手でチューニングした発音記号列（A）を作成する⁴
- (3) サンプルテキストを従来システムに入力し、発音記号列（B）を得る
- (4) サンプルテキストを本手法を実装したシステムに入力し、発音記号列（C）を得る
- (5) 得られた発音記号列（B）、（C）のポーズ挿入位置を、発音記号列（A）のポーズ挿入位置と比較する

4.3 評価方法

評価は、予備調査時の評価方法と同様に、ポーズ挿入正解率、ポーズ挿入精度向上率、ポーズ挿入精度降下率という 3 つの観点で行なった。

4.4 評価結果

評価結果を表 5 に、更に、ポーズ挿入精度向上例を発音記号サンプルを挙げて示す。

表 5 評価結果

ポーズ挿入正解率（従来システム）	80.7%
ポーズ挿入正解率（本手法実装システム）	95.5%
ポーズ挿入精度向上率	77.8%
ポーズ挿入精度降下率	7.7%

- ひとまとまりの意味を成す連文節中に目立ったポーズが挿入されるなどの致命的なポーズ誤りに対するポーズ挿入精度向上率は 100%
- ポーズ挿入精度降下率は係り受け解析誤りを含む
- QJP の評価文に対する構文解析精度（係り受け正解率）は 86.6%

4 2 人の人間がチューニングし、意見が分かれた箇所は一緒に実際に出力音声聞き比べて検討し、一意に意見をまとめた

【ポーズ挿入精度向上例】⁵

●入力テキスト

「その女性は赤い長めのオーバーコートを着て、…」

●従来システムが出力した発音記号列

ソノ/ジョセーワ/アカイ////ナガメノ/オーバーコ'ートオ//キテ////

「赤い」と「長め」の間に長いポーズが挿入されているが、この発音記号を音声出力すると、文意と異なった意味にとれるほど不自然である。隣接文節との係り受けの可否で検査するため、「女性は」と「赤い」の係り受けの関係が強く、「赤い」と「長め」の係り受けの関係が弱いことにより、こういった失敗が生じていた。

●本システムが出力した発音記号列

ソノ/ジョセーワ////アカイ//ナガメノ/オーバーコ'ートオ/キテ////

本手法による発音記号では、文意に即した自然な音声出力が得られた。遠い係り先まで検査するため、「女性は」の係り先可能性として、「赤い」と「着て、」を検出する。QJPにより「着て」との係り受けが尤もらしいと判断され、係り受け文節間の距離4で「女性は」の後に長いポーズが挿入される。一方、「赤い」は「オーバーコートを」に係るので、係り受け文節間の距離2で「赤い」の後には「女性は」の直後のポーズより短いポーズが挿入される（図2参照）。

【係り受け木構造】

[1]: 「その
[2]: 「女性は
[3]: 「赤い
[4]: 「長めの
[5]: 「オーバーコートを
[6]: 「着て、
:

【係り受け関係】

<連体詞連体:体言句>
<主格主題:動詞句>
<連体形連体:体言句>
<の連体:体言句>
<を連用:動詞句>
<て連用:動詞句>
:

図2 QJPの出力した係り受け

⁵ 発音記号列における「/」はアクセント句境界またはポーズ挿入位置を示す。また、「/」の数が1のときはアクセント句境界を表し、2以上のときはポーズを表す。2以上の場合、数が多いほどポーズ挿入尤度が高い（＝ポーズの長さが長い）ことを示す。以降の例についても同様である。

4.5 考察

評価結果より、文節数に制限のない係り受け文節間の距離、および係り受け関係をポーズ挿入処理に利用することが、ポーズ挿入精度の向上に寄与することが確認された。

一方、適切なポーズが挿入できない原因を調査すると、約54%が構文解析処理の誤りに起因するものであった。その他、係り先が遠い文節が続いた場合のポーズ長の調整不足によるもの、規則2の係り受け関係の分類の甘さによりポーズ挿入尤度が効いていないものがあった。

4.5.1 構文解析精度との関係

本システムにおいては、規則1により、係り先文節への距離5以上の場合、挿入尤度が同じ(大ポーズレベル)になるため、距離5の係り受けを、距離10の係り受けと誤っても、係り受け誤りは吸収されるため、構文解析誤り全てがポーズ挿入処理に悪影響を及ぼすものではない。逆に距離4以下の係り受けに対しては精度の高さが求められる。構文解析処理の誤りに起因するポーズ挿入誤りは、QJPの係り受け解析規則をより高精度にすることで、減少することが期待できる。

4.5.2 ポーズ長の調整

接続詞句や副詞句など、係り先が遠く独立性の高い文節の直後には、規則に基づく長いポーズが挿入されるが、これらの句が連続して続く場合には短いモーラ長の句の間に長いポーズが続けて挿入される。構文的区切り位置であることは明らかに分かるが、自然な読み上げには聞こえない。

下記の例では、「しかし」、「仮に」、「それが」の係り先文節への距離(d)が4以上であるため、各文節末に長いレベルのポーズが挿入され、音声出力した場合、不自然に聞こえてしまう。

【長いポーズ連続挿入例】

●入力テキスト

「しかし、仮にそれが実証に一番都合のいい方法であるにしても・・・」

●本システムで出力される発音記号列

シカ' シ/////カリニ/////ソレガ/////ジッショーニ///イチバン//ツゴーノ/
イ' ー/ホーホーデ/ア' ルニ/シテ' モ...

文法的要因の他に境界前後の句のモーラ長がポーズ位置やポーズ長に影響を与えるということとは言及があり(箱田, 佐藤 1980; Tsukada 1996, など)、モーラ長も無視できない要因である。上記例のように3モーラ程度と短く、独立性の高い文節間に長いポーズが連続して挿入される

場合には、ポーズ挿入尤度の低い境界におけるポーズのポーズ長レベルの格下げなど、2.5節のポーズ挿入処理のアルゴリズムの6の逆を実施するなど、呼気段落に基づくモーラ長の最短閾値の設定で対処していくことが考えられる。

4.5.3 係り受け関係の分類

本手法ではQ J P構文解析系の係り受けの関係を大まかに6種に大分類してポーズ挿入尤度を持たせたが、尤度が効いていない事例があった。ガ格やヲ格など格関係を別分類にする、また、係り受け関係の出現順を考慮するなど、まだ再考の余地があると考えられる。

5 おわりに

テキスト音声合成のポーズ挿入処理に軽量・高速な構文解析処理を導入し、一文全体の係り受け情報を利用して構文的区切り位置の同定、同位置へのポーズ挿入、最適なポーズ長の実現を実システム上で試みた。その結果、以下の2点を確認することが出来た。

- 一文全体の係り受け解析処理を導入した本手法のポーズ挿入処理方法の方が、隣接間係り受け処理を用いたポーズ挿入処理よりも挿入精度が高く有効である
- 係り受け文節間の距離と係り受け関係の情報以外に、同レベルポーズ間のモーラ長閾値の調整が必要である。さらに、係り受け関係の細分や規則としての適用順を考慮することでポーズ挿入精度の向上が期待できる

本稿では、文節数に制限のない係り受け処理をテキスト音声合成システムの言語処理部に実装して、ポーズ挿入精度の向上効果を確認した。今後は、この結果を反映した規則音声合成システムを実用レベルにするために、さらに大量のテキストを利用してデータやシステムのチューニングを進める予定である。

また、構文解析処理の精度を高めると共に、その導入により得られる情報をもっと有効活用し、テキスト読み上げだけでなく、ユーザの任意な出力形態要求に対応できる出力形式について検討していく予定である。

謝辞

本稿作成の過程で適切な助言をくださった(株)リコー情報通信研究室第34研究室的藤本潤一郎室長に感謝いたします。

参考文献

- Doi, S., Iwata, K., Murakai, K., and Mitome, Y. (1994). "Pause Control in Japanese Text-To-Speech Conversion System with Lexical Discourse Grammar." In *Proceedings of the ICSLP*, pp. 743-746.
- Iwata, K., Mitome, Y., and Watanabe, T. (1990). "Pause Rule for Japanese Text-To-Speech Conversion using Pause Insertion Probability." In *Proceedings of the ICSLP*, pp. 837-840.
- Kameda, M. (1996). "A Portable and Quick Japanese Parser : QJP." In *Proceedings of the COLING*, pp. 616-621.
- Tsukada, H. (1996). "A left-to-right processing model of pausing in Japanese based on limited syntactic information." In *Proceedings of the ICSLP*.
- 海木延佳, 句坂芳典 (1996). "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, **J79-D** (9), 1455-1463.
- 望主雅子, 亀田雅之 (1998). "品詞体系の変換を考慮した中間品詞体系と品詞変換表の設計." 言語処理学会第4回年次大会講演論文集, pp. 684-687.
- 亀田雅之 (1996). "軽量・高速な簡易日本語解析系QJP." テクニカル・レポート 22.
- 浅野久子, 松岡浩司, 市井亮美, 大山芳史 (1995). "テキスト音声変換における読み・韻律付与処理の評価～ニュース文を対象として～." 情報処理学会第51回全国大会講演論文集2, pp. 109-110.
- 張玉潔, 尾関和彦 (1997). "文節間係り受け距離の統計的性質を用いた日本語文の係り受け解析." 自然言語処理, **4** (2), 3-19.
- 潤淵謙一, 濱上知樹, 古村光夫 (1996). "文節の最終形態素に着目した韻律記号付与規則." 日本音響学会講演論文集, 1-P-7号, pp. 295-296.
- 塚田元, 田中一男 (1996). "高品質音声合成用テキスト解析技術." テクニカル・レポート 10.
- 丸山宏, 荻野紫穂 (1992). "日本語における文節間係り受け関係の統計的性質." 情報処理学会第45回全国大会講演論文集3, pp. 173-174.
- 村上春樹 (1995). 国境の南、太陽の西 (文庫版). 講談社.
- 宮崎正弘, 大山芳史 (1986). "日本文音声出力システムの言語処理." テクニカル・レポート 2.
- 佐藤奈穂子 (1997). "簡易日本語解析系QJPを用いたポーズ位置設定." 情報処理学会第55回全国大会講演論文集2, pp. 17-18.
- 日本電子工業振興協会 (1995). オフィスオートメーション(OA)機器の標準化に関する調査研究報告書. 日本電子工業振興協会.
- 杉藤美代子 (1988). 談話におけるポーズとイントネーション. 明治書院.
- 藤尾茂, 句坂芳典, 樋口宜男 (1997). "確率文脈自由文法を用いた韻律句境界とポーズ位置の予

- 測.” 電子情報通信学会論文誌, **J80-D-2** (1).
- 箱田和雄, 中畠信弥, 広川智久 (1989). “文章音声の音調結合型導出規則の検討.” 電子情報通信学会技術報告, **SP89** (5).
- 箱田和雄, 佐藤大和 (1980). “文音声合成における音調規則.” 電子情報通信学会論文誌, **J63-D** (9), 715-722.
- 鈴木和洋, 斎藤隆 (1995). “日本語テキスト音声合成のためのN文節構造解析とそれに基づく韻律制御.” 電子情報通信学会論文誌, **J78-D-II** (2), 177-187.

略歴

- 佐藤 奈穂子**: 1990年東京女子大学文理学部日本文学科卒業。同年, (株)リコー入社。日本語処理, 音声言語情報処理の研究開発に従事。言語処理学会会員
- 小島 裕一**: 1989年早稲田大学大学院理工学研究科修士課程卒業。同年, (株)リコー入社。音声言語情報処理の研究開発に従事。
- 望主 雅子**: 1986年東京女子大学文理学部日本文学科卒業。同年, (株)リコー入社。日本語処理, 音声言語情報処理の研究開発に従事。計量国語学会, 情報処理学会, 言語処理学会各会員。
- 亀田 雅之**: 1977年東京大学教養学部基礎科学科卒業。79年同大学院理学系研究科相關理化学専門課程[化学物理]修士修了。同年, 富士通(株)入社後, FHL出向を経て, 82年より富士通研究所にて自然言語理解, 知識表現, 機械翻訳の研究開発に従事。87年本田技術研究所入社, 和光研究センター勤務。88年リコー入社, 現在に至る。自然言語処理(特に日本語解析とその応用)の研究開発に従事。情報処理学会, 言語処理学会各会員。

(1998年4月3日 受付)

(1998年7月27日 採録)