

# 係り受け構造を用いたポーズ位置およびポーズ長の決定

清水 司      梅村 祥之      原田 義久

(株) 豊田中央研究所 機械認識研究室

## 1 はじめに

現在、自動車を取りまく情報環境が変わりつつある。ドライバは、CD-ROMなどの車載情報以外にもネットワークを通じて、交通情報、タウン情報、電子メールなど様々な情報にアクセスできるようになってきた。それに伴い、自動車内における音声合成のニーズが高まっている。

しかし、現在の合成音声の品質は、運転中に気軽に聞けるほど十分なものではない。自然で聞き取り易い合成音声を得るための1つとして、適切な位置および長さのポーズを挿入すること（ポーズ制御）が挙げられる。

ポーズ制御に関して、文献[1]では、10人分の音声データから局所的な構文構造とポーズ位置の関係を調べ、特定の構文構造を持つ場合にポーズが挿入される易いことが報告されている。文献[2]では、局所的な（3、4文節の範囲）係り受け解析を行い、得られた係り受けパターンに基づいてポーズ制御を行っている。しかし、局所的な係り受け解析では、大局的な構造をとらえることができず、構文的まとまりの1節中にポーズが挿入され、不自然な読み上げになるとの指摘がある[3]。

最近、1文全体の係り受け解析から求めた文節間の係り受け距離によって、ポーズ位置およびポーズ長を制御することの有効性が報告されている[3]。文献[3]では、文中の各文節末に、係り受け距離に基づいて、ポーズが挿入されやすいかどうかの指標をポーズ挿入尤度として設定し、ポーズ挿入尤度に従って適当なポーズを挿入している。また、係り受け距離が1（直後の文節に係る）の場合には、係り受けの関係に基づいたポーズ挿入尤度を併用している。

本稿では、文節間の係り受け距離によるポーズ挿入尤度と文節属性（主に従属節のタイプ）によるポーズ挿入尤度を併用するポーズ制御手法の妥当性の検討を行った。

## 2 文節属性によるポーズ制御手法

### 2. 1 係り受け距離によるポーズ制御

例文（1）では、文節「倒れるので」の直後に最も長いポーズを挿入するのが適切であると考えられる。

例文（1）「不況で 中小企業が どんどん 倒れるので 競争相手が 多くなった。」

しかし、係り受け距離に基づくポーズ制御では、文節「倒れるので」の係り受け距離2に対して、文節「不況で」、文節「中小企業が」の係り受け距離がそれぞれ、3、2となり、適切なポーズ制御ができない。そこで、次節で述べる文節属性によるポーズ挿入尤度を導入した。

### 2. 2 文節属性

本稿で提案する手法では、ポーズの位置および長さを決定するために、文節属性と呼ぶものを定義した。文節属性は主に従属節のタイプに基づいて分類し、それぞれの文節属性の種類に対して、表1に示すポーズ挿入尤度を与えた。

### 2. 3 本手法の概要

#### （1）形態素解析および係り受け解析

まず、テキストの形態素解析処理を行う。形態素解析システムには、奈良先端科学技術大学院大学で開発されたChaSen Ver. 1.51[4]を用いた。次に、係り受け解析処理を行う。係り受け解析システムは、著者が開発したものであり、ChaSenの品詞体系に基づく形態素情報ベースの規則による係り受け解析システムである。前節で説明した文節属性は、係り受け解析処理の文節化の段階で、形態素情報ベースの規則によって付与されるようになっている。

#### （2）係り受け距離によるポーズ挿入尤度

係り受け解析結果から各文節の係り受け距離を求め、係り受け距離によるポーズ挿入尤度を式（1）で表されるシグモイド関数によって定めた。ここで、 $d$ は係り受け距離である。

表1 文節属性とそのポーズ挿入尤度

文節属性	ポーズ挿入尤度	係り受け距離が1の場合の係数
逆接	1.0	0.5
原因理由	1.0	0.5
目的	1.0	0.5
条件譲歩	0.857	0.5
付帯状況様態	0.857	0.5
並列	0.714	0.0
形式名詞並列	0.714	0.5
副詞	0.571	0.5
接続詞	0.571	0.5
時表現	0.571	0.5
提題	0.429	0.5
テ形提題	0.429	0.5
引用	0.289	0.5
疑問	0.289	0.5
形式名詞	0.289	0.5
NIL (上記以外のもの)	0.143	1.0

$$d\_score(d) = \frac{1}{1 + e^{-(d-3)}} \quad (1)$$

シグモイド関数を用いた理由は、係り受け距離が大きくなっても若干の尤度差を生じさせるためである。式(1)を用いることで係り受け距離によるポーズ挿入尤度は0から1の範囲である。

### (3) 文節属性によるポーズ挿入尤度

係り受け解析結果から得られた文節属性に対して、表1に示すポーズ挿入尤度を与えた。今回、ポーズ挿入尤度は、文節属性の種類に応じて7段階に設定し、係り受け距離によるポーズ挿入尤度との整合性を持たせるため0から1の範囲とした。

### (4) ポーズ挿入尤度によるポーズの決定

今回は、(2)、(3)で求めた2種類のポーズ挿入尤度を単に足し合わせたものを最終的なポーズ挿入尤度とした。ただし、予備検討により係り受け距離が1の場合は、文節属性によるポーズ挿入尤度の影響が大きくなり悪影響を及ぼすことが分かった

ので、この場合に限り、文節属性によるポーズ挿入尤度に表1で示す係数を乗じた。

最終的なポーズ挿入尤度の高い文節境界から順に、適当なレベルのポーズを挿入する。

## 4 評価

### 4.1 評価の対象文

本手法を評価するために、ATR音声データベース(話者MYI)の502文<sup>1</sup>を用いた。このデータベースには、1人分の話者のポーズ位置および長さがテキスト情報として含まれている。ポーズ挿入尤度の評価では、この話者のポーズ位置および長さを基準とした。

### 4.2 ポーズ挿入尤度の評価

1文中の各文節境界に対して、ポーズ挿入尤度を求め、それらの中で最大のポーズ挿入尤度に着目して評価した。最大ポーズ挿入尤度を持つ文節境界に対して、ATR音声データベース(話者MYI)でのポーズの有無(ポーズが有る場合は、1文中におけるポーズ長の順位)の頻度を調べた。提案手法との比較のために、ポーズ挿入尤度として、係り受け距離のみによるもの、文節属性のみによるものについても評価を行った。

結果を表2に示す。最大ポーズ挿入尤度を持つ文節境界が複数あった文については、評価の対象から除外した。係り受け距離のみによるポーズ挿入尤度を用いた場合、最大ポーズ挿入尤度の文節境界が最大ポーズ長を持つ文節境界に一致したのは、246境界であった。これに対して、文節属性によるポーズ挿入尤度を併用することによって、37境界増えた。例文(1)は表3に示すように改善され、適切な位置(文節「倒れるので」の直後)に最長のポーズを挿入することができた。また、最大ポーズ挿入尤度の文節境界が2番目に長いポーズ長を持つ文節境界に一致したのは、6境界増えた。一方、最大ポーズ挿入尤度を持つ文節境界に対して、ポーズが挿入されていない数は、文節属性によるポーズ挿入尤度を併用することによって10境界しか増えていない。

これらの結果から、係り受け距離によるポーズ挿入尤度に文節属性によるポーズ挿入尤度を併用する

<sup>1</sup>ATRの音声データベースには503文含まれているが、係り受け解析ができなかった文が1文あり、評価対象から除外した。

表2 最大ポーズ挿入尤度とATR音声データベースのポーズの関係

	最大の ポーズ	2 番目に 長いポーズ	3 番目以降に 長いポーズ	ポーズ無	複数の最大 尤度を持つ文	ATRデータベ ースでは文節境界 ではない
係り受け距離のみ	246	28	4	130	79	15
文節属性のみ	234	32	1	111	106	18
係り受け距離 + 文節属性	283	34	4	140	24	17

表3 例文(1)の各文節境界でのポーズ挿入尤度(カッコ内は尤度の順位)

不況で 中小企業が どんどん 倒れるので 競争相手が 多くなった。					
ポーズ	pause 270				
文節属性	NIL	NIL	NIL	原因理由	NIL
係り受け距離のみ	0.5 (1)	0.27 (2)	0.12 (3)	0.27 (2)	0.12 (3)
文節属性併用	0.64 (2)	0.41 (3)	0.26 (4)	1.12 (1)	0.26 (4)

ことによって、ポーズ制御の精度が向上することが示された。

## 5 考察

### 5.1 文節属性によるポーズ挿入尤度

今回用いた文節属性のポーズ挿入尤度の順位は経験的に決めたのであるが、その値は理論的、経験的な根拠に基づいているものではない。文節属性間のポーズ挿入尤度の相対的な大きさは、音声データコーパスから統計的に求めることが考えられる。今回、ポーズ挿入尤度の絶対的な大きさは、文節間距離によるポーズ挿入尤度との足し合わせ時の整合をとるつもりで0から1の値を用いたが、整合がとれているかは十分な検討が必要である。また、2種類のポーズ挿入尤度を足し合わせるときに、重みを1:1としたが、重みの最適値は実験的に求める必要がある。

### 5.2 形態素解析および係り受け解析の精度

表2に示した結果は、形態素解析および係り受け解析の誤りを含んだ結果である。係り受け解析の正解率は、評価に用いた502文に対して文節単位で79.8%であった。係り受け解析の精度が向上すれば、

係り受け距離によるポーズ挿入尤度の精度が向上することが考えられ、ポーズ制御の向上につながるであろう。一方、文節属性によるポーズ挿入尤度は、係り受け解析の精度にあまり関係がないと考えられる。したがって、今回の結果は、文節属性を導入することによる直接的な効果ではなく、文節属性が構文解析結果を補正した結果と考えることができるかもしれない。これについては、今後、係り受け解析結果を補正して同様な評価を行い、検討する必要がある。

### 5.3 読点情報の効果

今回、評価に用いた文は、読点が全く入っていないものであった。従来のポーズ制御の多くの研究では、読点に対してある一定のポーズ長を対応させていた。しかし、読点の挿入は、文を書く人の個人差が大きい。したがって、読点を直接的に用いてポーズ制御を行った場合、読み上げ対象文を誰が書いたかで、合成音声の品質が左右され、一定したものにならないことが考えられる。読点を直接的に用いることには注意が必要であろう。今回の結果は、文節属性を導入することによって、読点の有無に関わら

ず、一定したポーズ制御が可能であることを示している。

## 6 まとめ

本稿では、音声合成においてポーズの制御を行うために、文節間の係り受け距離によるポーズ挿入尤度だけではなく、文節属性によるポーズ挿入尤度も併用する手法を提案し、ATR音声データベースを用いて評価した。その結果、文節属性によるポーズ挿入尤度の決定方法などに検討すべき課題が残るものの、手法としての枠組みの有効性を示した。

## 参考文献

- [1] 海木延佳、匂坂芳典：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 D-II、Vol. J79-D-II、No. 2、pp. 1455-1463 (1996)
- [2] 鈴木和洋、斉藤隆：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 D-II、Vol. J78-D-II、No. 2、pp. 177-187 (1995)
- [3] 佐藤奈穂子、小島裕一、望主雅子、亀田雅之：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、自然言語処理、Vol. 6、No. 2、pp. 117-132 (1999)
- [4] <http://cactus.aist-nara.ac.jp/lab/nlt/chasen.html>