

## 韻律構造を考慮した対話音声の発話速度の分析と合成

川波弘道, 広瀬啓吉

東京大学工学系研究科電子情報工学専攻

〒113-8635 東京都文京区本郷 7-3-1

Tel.: (03) 3812-2111 (ext:6767,6787)

E-mail: {kawanami,hirose}@gavo.t.u-tokyo.ac.jp

あらまし 対話音声合成の韻律規則作成のために、対話調音声と朗読調音声の比較研究を進めている。すでに発話速度の合成規則も作成しているが、これは一韻律文に対して一律の持続時間制御を行うものであった。ここでは朗読音声との比較による対話音声のモーラ持続時間短縮率を中心に、発話速度と基本周波数パターン ( $F_0$  パターン) により定義される韻律構造との対応を調べた。具体的には、 $F_0$  パターン重畳モデルに基づく韻律文、韻律節、韻律句、韻律語の4つの韻律単位と発話速度との関係を分析した。韻律文では、対話音声の発話速度は、文頭で朗読音声程度の発話速度であり、それが徐々に大きくなった後再び文末に向けて減少するが、同様の傾向が、程度は小さくなるものの下位レベルでも見られることが分かった。その結果に基づき、韻律構造を反映した発話速度制御規則に対する考察を行った。

キーワード 対話音声, 発話速度, 韻律構造, 音声分析, 音声合成

## Analysis and Synthesis of Speech Rate in Dialogue Speech Based on Prosodic Structures

Hiromichi Kawanami, Keikichi Hirose

Department of Information and Communication Engineering,  
School of Engineering, University of Tokyo

7-3-1 Hongo, Bunkyo-ku, Tokyo 113, Japan

Tel.: 03 3812 2111 (ext:6767,6787)

E-mail: {kawanami,hirose}@gavo.t.u-tokyo.ac.jp

**Abstract** We have been conducting a comparative study on the dialogue-style and reading-style, to construct prosodic rules for the synthesis of dialogue speech. In the current report, analyses were conducted for the speech rate and rules were constructed with respect to the prosodic structure defined based on the  $F_0$  contours. We have formerly defined 4 levels in prosodic units based on the  $F_0$  contour modeling (so called Fujisaki's modeling); prosodic sentence, prosodic clause, prosodic phrase and prosodic word. These units are adopted for the current analysis. In a prosodic sentence, the speech rate of dialogue-style speech started with a value close to the case of reading-style, gradually increased, and then slowed down toward the end of the sentence. Similar tendency was also observed in lower level units, but the degree of speech rate change in a unit was smaller for minor levels. Based on the result, we give consideration to the prosodic rules for synthesizing dialogue-style speech rate reflecting their prosodic structure.

**Keywords** dialogue speech, speech rate, prosodic structure, speech analysis, speech synthesis

## 1 はじめに

対話システムの応答生成部の合成器を想定した場合、必要な言語情報を伝達するのみならず、音声言語に特有なパラ・非言語情報も伝達可能で、かつ人間にとって自然に感じられる音声であることが期待される。それをふまえて、これまで人間らしい自然な対話調の韻律の音声を合成する手法の開発をすすめている。

すでに、対話音声を対象として、重畳モデルと数量化手法に基づく基本周波数パターン ( $F_0$  パターン) の制御要因の分析、朗読音声との比較による発話速度 (モーラ持続時間) の分析を行い、対話音声合成のための韻律制御規則を作成した [1]。ただし発話速度に関しては、1文ごとにその変化の概略を一律に制御するのみであった。そこで、文の構造が音声において最も明確に表れるのは  $F_0$  パターンであるという立場から、発声の単位を定義し、それに基づいた発話速度の分析を行った [2]。

ここでは  $F_0$  パターンで表現される韻律の単位として韻律語から韻律文までを考え、また分析対象資料数を増大させることで、発話速度と韻律的文構造の関係をより詳しく調査した。そして、その結果に基づき、より適切な発話速度制御の規則化に対する考察を行った。

## 2 韻律の構成

発話速度は音声言語に特有な特徴であり、従ってその文構造との関係を分析するに際し、文法的構造よりも発声に表れる構造に着目した方が見通しのよい結果が得られると考えられる。音声言語におけるこの様な韻律構造は、 $F_0$  パターンによって明確に表現されることが知られている [3]。そこで定義された韻律単位のうち、韻律文と発話速度との関係についてはすでに報告している [4] が、ここでは韻律文も含めさらに、韻律節、韻律句および韻律語と発話速度との関係を分析する。

一般的に文音声では、文法上の文が韻律文として、連節・節が韻律節として、ICRLB (Immediate Constituent with Recursively Left-Branching structure, 右枝分かれ境界で前後を区切られ、かつ左枝分かれ境界のみを含む単語連鎖) が韻律句として韻律的に表現されるが、発声機構の生理的制約で韻律節・韻律句のモーラ数が制限されるなど、統語構造と韻律上の構造は一對一に対応するとは限らない。

そこで、発話速度と文構造との関連を考える際に、韻律上の構造を示す各々の単位を  $F_0$  パターンの観点から次のように定義し、これらを単位として発話速度変化を分析する。

韻律文 (prosodic sentence) : 比較的長い休止に挟まれ、立て直しのフレーズ成分に対応する区間。

韻律節 (prosodic clause) : 比較的短い休止に続く、追加のフレーズ成分に対応する区間。

韻律句 (prosodic phrase) : 追加のフレーズ成分に対応する区間。

韻律語 (prosodic word) : 1つのアクセント成分に対応する区間。

なお、下位の単位が上位の単位と一致する資料は上位の単位のみにも属するとして分析した。

## 3 音声資料と分析手法

演劇活動をしている東京方言男性話者 6 名、女性話者 4 名が、あらかじめ用意したテキストに基づいて行った模擬対話を収録し、これを対話資料として用いた。対話は 54 文と 70 文からなる 2 セットで、ともにクライアントがエージェントのアドバイスを受けてスキープランを作成するという内容である。比較のために同一内容を朗読したものも収録し、朗読音声資料とした。

今回は、特に対話音声としての評価が高い、男性話者 SH の音声を分析した。次に示すように発話速度と韻律構造との関連を対話音声と朗読音声との比較において調査する必要性から、ここでは両音声において韻律構造が同一であるものを分析に用いた。

発話速度の分析は、対話音声の各モーラの持続時間を朗読音声のそれと比較したモーラ持続時間短縮率 (Reduction rate of mora duration) なるパラメータを以下のように定義して行った。

$$\text{Reduction rate} = \frac{dur_r - dur_d}{dur_r} \quad (1)$$

ただし、

$dur_r$  : 朗読音声でのモーラ持続時間、

$dur_d$  : 対話音声でのモーラ持続時間。

これは、モーラ持続時間自体は構成音素等によって変化するため、それを極力押えた上で上位の構造 (韻律構造) と発話速度との関係を明確にしようとする意図である。なお、両音声のモーラ持続時間の絶対値も必要に応じて参照した。

モーラ持続時間は、音声波形から次の様にして測定した。まず、発声内容を音素表記したテキストファイルを与え、HTK (Ver.2.1) を用いた強制的なアラインメントを行って音声波形を音素単位に区分する。次に視察による修正を加えてモーラ持続時間を計測する。この際、2モーラ音素である長母音については、便宜的に (子音部も含め) 音素長を等分してモーラの持続時間とした。このようにしてモーラ持続時間を対話音声と朗読音声の対について求め、(1) 式によりモーラ持続時間短縮率を計算した。

Table 1. 各単位の総数とモーラ数の内訳

	総数	モーラ数																
		2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18以上
韻律文	63	13	2	1	5	2	6	4	6	3	4	2	2	0	4	1	1	7
韻律節	7	2	0	1	0	1	0	0	0	0	1	1	0	0	0	0	1	
韻律句	38	2	3	4	5	3	6	3	5	2	3	1	1	0	0	0	0	
韻律語	111	17	12	16	13	13	11	6	8	1	0	1	0	0	0	0	0	

Table 2. 韻律語の各モーラ数のアクセント型の内訳

総数	モーラ数											
	2	3	4	5	6	7	8	9	10	11	12	
0型	1	2	6	1	1	0	0	0	0	0	0	
1型	16	2	8	2	2	1	0	0	0	0	0	
2型	-	8	1	3	3	1	0	0	0	0	0	
3型	-	-	1	4	3	0	0	0	0	0	1	
4型	-	-	-	3	3	2	0	0	0	0	0	
5型	-	-	-	0	3	0	1	0	0	0	0	
6型	-	-	-	-	1	5	2	0	0	0	0	
7型	-	-	-	-	-	1	4	0	0	0	0	
8型	-	-	-	-	-	-	1	0	0	0	0	
9型	-	-	-	-	-	-	-	1	0	0	0	
10型	-	-	-	-	-	-	-	-	0	0	0	
11型	-	-	-	-	-	-	-	-	-	0	0	
12型	-	-	-	-	-	-	-	-	-	-	0	

さらに各韻律単位において、その構成モーラ数で分類した上でそれぞれ、単位内の短縮率変化の平均と標準偏差を計算した。ただし、韻律語の場合はさらにそのアクセント型(0型,1型,2型,..)で分類した上で計算をした。

分析に用いた韻律文、韻律節、韻律語、韻律語の総数およびその内訳を Table 1 に、韻律語のモーラ数別分類の内訳を Table 2 に示す。韻律構造は、F<sub>0</sub> モデルを用いた Analysis-by-Synthesis に基づいて決定した。

## 4 分析結果

韻律単位別に、短縮率の値とその変化パターンについて見られた特徴を示す。ただし、2モーラのものには「はい」、「では」などであり、3モーラ以上のものとは別の扱いとした。

### 4.1 モーラ持続時間短縮率の一般的特徴

#### 4.1.1 韻律文内での傾向

韻律文内での発話速度変化の特徴として、次の3点の特徴があることが既に明らかとなっている [4]。

- (1) 発話速度が増加する。
- (2) 短縮率は、区間の頭では朗読音声なみで、それから半ばすぎにかけて増加し、その後ふたたび区間末にかけて穏やかに低下する。
- (3) 休止の前のモーラの持続時間は著しく延長されることがある。(phrase final lengthning)

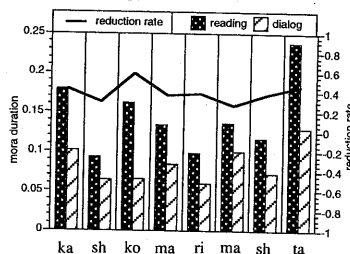


Fig. 1. 「かしこまりました」の短縮率変化(話者 SH).

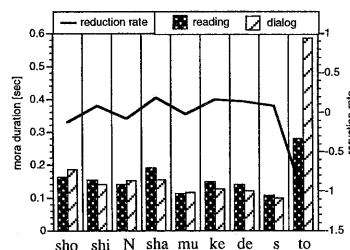


Fig. 2. 「初心者向けです」の短縮率変化(話者 SH).

(1), (2) については殆どの資料で見ることができるが、(2)の傾向が見られない例 (Fig.1) もあった。(3)の例を Fig.2 に示す。

#### 4.1.2 韻律節内および韻律節間での傾向

分析資料中には朗読音声との比較が可能な対話音声の韻律節は少数しかなかったが、それらでは韻律文における特徴 (1), (2) が見られた。その例を Fig.3 に示す。

Fig.4 の例では、後半の韻律節「そうです」が高い短縮率で始まっている。これは、「そう」の2重母音(oo)が対話音声で大きく短縮されるという特徴を示すものと考えられる。

#### 4.1.3 韻律句

韻律句に関しても (1), (2) の傾向が多く見られた。ただし、韻律句境界付近での短縮率の低下は韻律文・韻

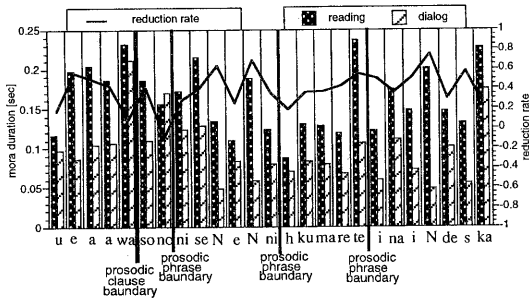


Fig. 3. 「ウエアーはその二千円に含まれていないんですか」の短縮率変化(話者 SH).

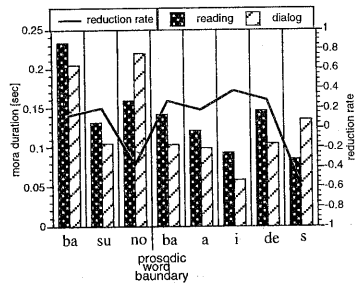


Fig. 6. 「バスの場合です」の短縮率変化(話者 SH).

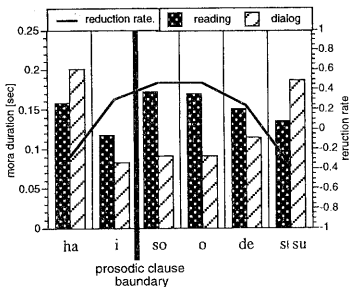


Fig. 4. 「はい、そうです」の短縮率変化(話者 SH).

律節の場合と比較して顕著ではない。韻律句はモーラ数が10モーラ以下程度のものが多く、その中に含まれる韻律語の種類の影響も大きい。

(2)と(3)の特徴が表れている例としてFig.5を示す。

#### 4.1.4 韻律語

全体的な傾向というのは見出しにくい。短縮率が変化する様子が  $F_0$  パターンのそれとの関連があると考えると、韻律語のアクセント型によって短縮率変化が様々な形で表れているのではないかと考えられる。また、構成モーラ数が5モーラ程度以下の小さいものが多

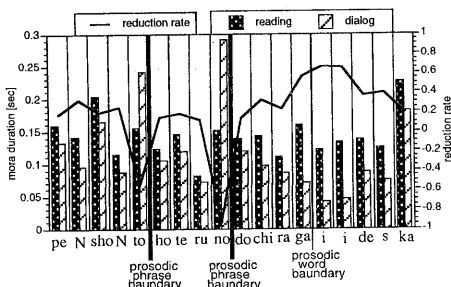


Fig. 5. 「ペンションとホテルのどちらがいいですか」の短縮率変化(話者 SH).

いため、音素構成の違いが短縮率変化パターンの傾向を見る際に障害となっていると思われる。

例えば、Fig.6は韻律語内で(2)の特徴が見られるが、音素に注目してみると「場合」の部分は母音が連続しているため、短縮率が高くなっているという可能性もある。

また同様に、前出のFig.5の「どちらが/いいですか」において、「がいい」の部分が高い短縮率を示しているのは韻律語境界にまたがる母音の連続部が極端に短縮された結果と思われる。

## 4.2 統計的処理結果

本節では、統計的な見地からの考察を加える。

2節で述べたように各単位について、韻律語についてはモーラ数とともにアクセント型別に分類して、それ以外はモーラ数のみの分類を行って平均値、標準偏差等を計算した。ここでは得られた結果の中から、サンプル数が5つ以上のものを取り上げ、検討を行う。

### 4.2.1 韻律文の平均的变化パターン

5モーラ(Fig.7)、7モーラ(Fig.8)の結果を示す。ともに前述の(1)、(2)の特徴が表われているのがわかる。

### 4.2.2 韻律句の平均的变化パターン

5モーラ(Fig.9)、7モーラ(Fig.10)の韻律句についての結果を示す。韻律文に比べて韻律構造は簡単なものとなるため、(1)、(2)の傾向がより顕著に表れている。

### 4.2.3 韻律語の平均的变化パターン

Fig.11, Fig.12に4モーラ1型アクセントの韻律語、4モーラ0型アクセントの韻律語についての分析結果、同様にFig.13, Fig.14に5モーラ1型アクセント、5モーラ0型アクセントの分析結果を示す。

これからは明確な特徴を見い出せないが、0型(平板型)アクセントでは末尾のモーラの短縮率の低下が1型よりも顕著ではないという傾向が示唆されている。

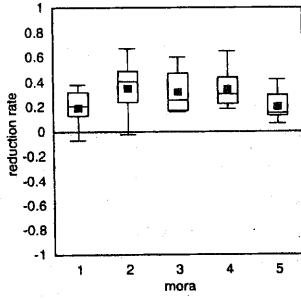


Fig. 7. 5 モーラ韻律文の分析結果.

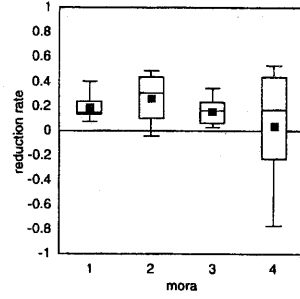


Fig. 11. 4 モーラ韻律語,1型アクセントの分析結果.

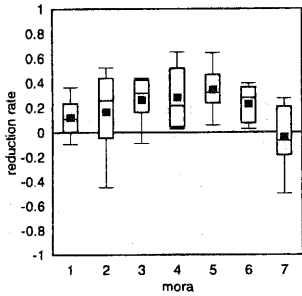


Fig. 8. 7 モーラ韻律文の分析結果.

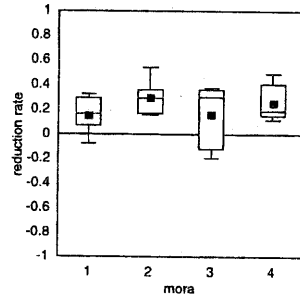


Fig. 12. 4 モーラ韻律語,0型アクセントの分析結果.

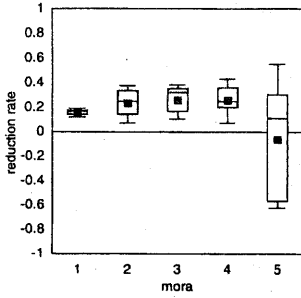


Fig. 9. 5 モーラ韻律句の分析結果.

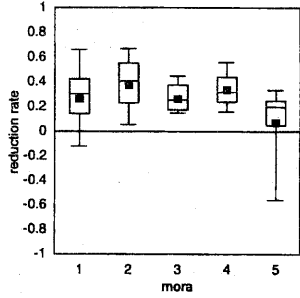


Fig. 13. 5 モーラ韻律語,1型アクセントの分析結果.

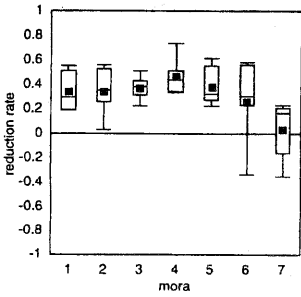


Fig. 10. 7 モーラ韻律句の分析結果.

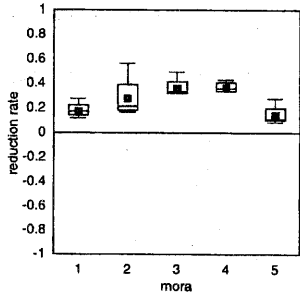


Fig. 14. 5 モーラ韻律語,0型アクセントの分析結果.

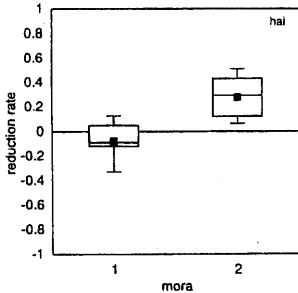


Fig. 15. 「はい」の分析結果.

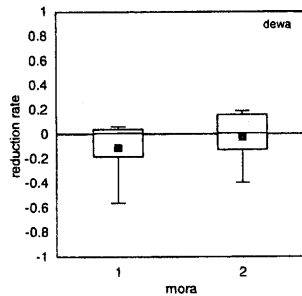


Fig. 16. 「では」の分析結果.

### 4.3 2モーラ韻律文

出現する2モーラ韻律文は回答である「はい」(9個)と、「では」(4個)である。ここでは韻律節「はい」「では」(各1個)も含めてその短縮率を考察する。「はい」, 「では」の平均的な短縮率変化パターンを Fig.15, Fig.16 に示す。

「はい」の場合も「では」の場合も、それぞれのグループの各音声の発話速度はグループ内で近い形を持つことがわかる。このことは、短い発声の発話速度制御を行う際には、韻律語に固有の発話速度制御規則を作成しておくことの必要性を示唆している。

## 5 制御規則の作成と音声合成の検討

これらの知見を考慮すると、発話速度制御には次のような規則を施すことが必要であると考えられる。

まず、朗読音声の発話速度制御を行った後、以下の対話音声用の制御を行う。

#### 1. 韻律構造に基づく発話制御

- (1) 韻律文に対する短縮率操作.
- (2) 韻律節に対する短縮率操作.
- (3) 韻律句に対する短縮率操作.
- (4) 韻律語に対して短縮率操作.

さらに、

#### 2. 音韻的な修正. (2重母音を対話においてさらに短縮

するなど)

#### 3. 必要に応じた韻律文末モーラ、韻律節末モーラの延長.

1.(1) ~ (3) はすべて基本的には各モーラに、標準的な単位内短縮率変化パターン(文頭から徐々に大きくして、そして文末に向けてまた次第に減少させる)を作成する処理である。変化パターンは単位あるいは構成モーラの数等で異なり、精密には大量データをもとにした統計的処理結果が必要である。1.(4)については、精密な制御のためには、さらに韻律語の  $F_0$  上昇・平坦・下降部での短縮率の違いを考慮する必要がある。2.,3. の具体的規則作成には更に分析が必要である。

## 6 おわりに

文の構造と発話速度との関係を、 $F_0$  パターンで定義される韻律節、韻律句、韻律語に着目して分析した結果を報告した。そしてその結果から、発話速度合成をどのような点から行うべきかを考察した。すでに、予備的な合成実験を行っているが、今後は実際に韻律規則を作成して本格的な合成実験を行い、評価する。

## 参考文献

- [1] 川波弘道, 広瀬啓吉, “対話音声の韻律的特徴の定量的分析による韻律規則の作成”, 音響学講論, pp.207-208 (1996-9).
- [2] 川波弘道, 広瀬啓吉, “対話音声における発話速度の分析と韻律規則の作成”, 音響学講論, pp.223-224 (1998-3).
- [3] Fujisaki, H., Hirose, K. and Takahashi, N., “Manifestation of linguistic information in the voice fundamental frequency contours of spoken Japanese”, *IEICE Trans. Fundamental of Electronics, Communications and Computer Sciences*, Vol.E76-A, No.11 pp.1919-1926 (1993-11).
- [4] Hirose, K., Sakata, M. and Kawanami, H., “Synthesizing dialogue speech of Japanese based on the quantitative analysis of prosodic features”, *Proc. 1996 International Conference on Spoken Language Processing, Philadelphia*, Vol.1, pp.378-381 (1996-10).