

TR-I-0172

品詞情報による音韻継続時間長の分析

*Hidehiro Inagaki Kaiki Nobuyosi  
and Yoshinori Sagisaka*

稲垣 英浩(\*) 海木 延佳 匂坂 芳典

1990.8

### 概要

多様な口調をもつ音声の合成を目標に、韻律制御規則の検討を行った。ここでは音韻継続時間長の制御、すなわち朗読調の文章における音韻継続時間長は、ある程度の精度で設定できるため、この時間長をもとに、通常の会話での口調における音韻継続時間の設定を考察し、誤差の分析を行った。その結果、誤差の最小値は、朗読調で26.84%、会話調で41.57%であった。おおまかに言って後者は前者の1.5倍の誤差が現れることが判った。

次にその誤差の著しく現れている部分について誤差を分析した。その結果、促音化している部分や、文末の強いニュアンスをもつ終助詞の影響がみられた。感動詞や逆接の接続詞の影響も予想したがはっきりとはあらわれなかった。

(\*)早稲田大学 この研究は著者がATRに夏期実習中に行ったものである。

ATR Interpreting Telephony Research Laboratories  
ATR 自動翻訳電話研究所

(\*)早稲田大学

## I はじめに

多様な口調をもつ音声の合成を目標に、韻律制御規則の検討を行った。  
ここでは音韻継続時間長の制御、すなわち朗読調の文章における音韻継続時間長は、ある程度の精度で設定できるため、この時間長をもとに、通常の会話での口調における音韻継続時間を設定することを考える。

## II 音声データベース

資料として、発声様式の異なる、朗読調、会話調、計7種類の発声様式のラベルファイルを用いた。それぞれが会話1、会話2、会話3、計3種類の会話例を含んでいる。朗読調のものとして、スピード通常、スピード速い、スピード遅いの計3種、会話調のものとして、親切、怒り、急ぎ、通常の4種を用いた。

## III 種々の発話様式における音韻継続長の制御

### 3. 1 種々の発話様式における音韻継続長の統計的性質

表1 にスピード通常、会話通常各発話様式について、各音韻、標本数、平均値、分散値、および各音韻種類毎の平均値を示した。また会話通常について、スピード通常を基準にした時の伸縮比を%表示した。

### 3. 2 種々の発話方式における音韻継続長の変換規則の検討

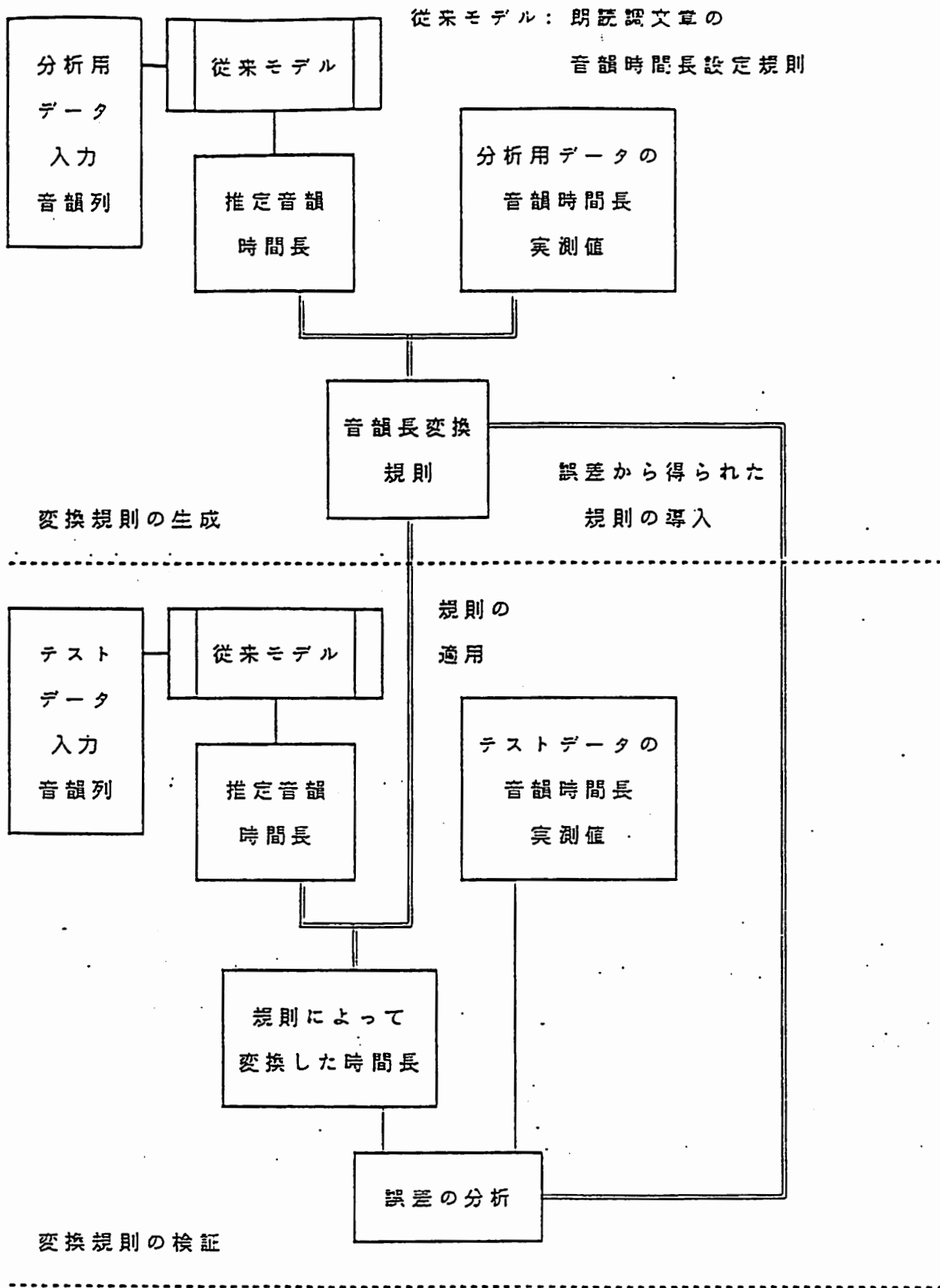


図1 音韻長変換規則の作成検証手続

すでに、音韻継続時間長を予測するモデルが試みられている。このモデルは、数量化1類を用いた線形モデルを利用している。音韻継続時間長を予測するために、制御要因を用いている

- ① 当該音韻
- ② 隣接前2音韻、後2音韻
- ③ 当該音韻の位置
- ④ 呼気段落のモーラ数
- ⑤ 促音化の有無
- ⑥ 母音部の品詞情報

分析データの音韻系列を従来モデル(朗読調の音韻時間設定を行うプログラム)に入力し、音韻時間長を推定する。品詞情報による分類を行い、品詞平均の単純誤差を加味して次に、推定値と実測値から簡単な変換規則を作る。一方、テストデータの推定値を従来モデルに入力し、音韻時間長を推定しておく。次に、このテストデータの推定値を分析データより得られた変換規則で変換した値と、テストデータの実測値との誤差を分析する。その分析のさいに品詞情報の側面から特に考察を行う。

このため、以下の3通りの単純な変換方法を用いて実験を行い、その推定誤差の分析を行った。

#### ① 全音韻長一律の変換

はじめに、モデルから予測された音韻継続時間長の全体の平均と実験値のそれとの比を用いる。この比を予測値に掛けて伸縮したもの、実験値との誤差を分析する。

#### ② 音韻種類毎の変換

次に、予測値と実験値の各音韻種類毎の平均値の比を用い、予測値を伸縮したものと、実験値との誤差を分析する。

ここで、音韻種類とは、母音(a, i, u, e, o)、無声摩擦(s, sh, h, f)、破擦( ts, ch )、無声破裂(p, t, k)、有声破裂(b, d, g)、鼻音(m, n)、有声摩擦(j, z)、半母音(w, y)、流音(r) で 母音は一括せずそれぞれの平均値をとる。

### ③ 各音韻毎の変換

予測値と実験値の各音韻毎の平均値の比を用い、予測値を伸縮したものと、実験値との誤差を分析する。

## 3. 3 各変換方式による結果

### ( 1 ) 全音韻長一律の変換における誤差

会話1、2、3 全てを用いて得られた規則を、それ自身に適用したときの誤差を表3 に示す。尚、表2は品詞分類で単純誤差を補正しない場合である。

会話1と3を用いて得られた規則を1、2、3各々に適用したときの誤差を表4に示す。

この結果から以下の事が判る。ただし、予測誤差の大きさを比べるのでは実測長の長いものほど大きく現れるので実測長平均で除したもので比較する。

・ 全体的にみて朗読調の誤差の約1.5倍、会話調で誤差が現れている。

実際、オープン実験では、

朗読通常が32.98%で最小だった。

また、急ぎ では49.85%で最大だった。

### ( 2 ) 音韻種類毎の変換における誤差

会話1と3を用いて得られた規則を1、2、3各々に適用したときの誤差を表5に示す。この結果から以下の事が判る

- ・ 全体的にみて朗読調の誤差の約1.5倍、会話調で誤差が現れている。  
実際、オープン実験では、  
朗読通常が28.50%で最小だった。  
また、急ぎ では46.82%で最大だった。

## ( 2 ) 各音韻毎の変換における誤差

会話1と3を用いて得られた規則を1、2、3各々に適用したときの誤差を表6に示す。この結果から以下の事が判る

- ・ 全体的にみて朗読調の誤差の約1.5倍、会話調で誤差が現れている。  
実際、オープン実験では、  
朗読通常が26.12%で最小だった。  
また、急ぎ では45.45%で最大だった。

## 3. 4 音韻毎の変換後の誤差の分析

従来モデルによって生成した規則だけでは誤差が一部大きく現れるところがあり、その個々の事象をとり、誤差の原因を分析刷る。各音韻ごとの変換の後の誤差の分布を図3 に示す。

この結果の一部、かなり値の大きい誤差例を紹介する。

呼びかけに使われている代名詞「これ」"e"	予測長	<	実測長
逆接の接続詞 「でも」 "o"	予測長	<	実測長
文末で意味を持つ終助詞 「です <u>ね</u> 」			

「ですか」「ますよ」「e””a””o”      予測長 < 実測長  
 融合化していない母音の連続部分「エアレーター」“a”  
 「カリフォルニア」“a”      実測長 < 予測長  
 撥音の前の母音「100えん」“e”      実測長 < 予測長  
 促音化している次の子音部分「きっぷ」“p”      予測長 < 実測長  
 長い単語に挟まれた並列助詞「酒や煙草」ではなく、  
 「スーツケースと航空かばん」“o”      予測長 < 実測長

これらのなかで多く見つかった誤差は、終助詞、促音化の部分である。  
 それ故、終助詞の母音部分の誤差平均長を表8 に、促音化部分の誤差平均長を表9  
 に示す。 図3 と併用してみるととくに終助詞の母音部分の誤差は顕著であ  
 る。

また、誤差がある句単位でバースト的に現れることを予測したのだが、その  
 ような傾向はあまり観察されなかった。 図2 に一つ前の音韻の誤差が -  
 30ms以下の時、-30ms~30msの時、30ms以上の時と分類して、当該音韻の誤差の  
 分布を表す。この結果をみても、誤差が連続して同じ方向に倒れるわけではな  
 い。

表の説明      表7までは二乗誤差、表8、9は単純誤差、但し負の数は予測長が  
 実測長より長いことを示す。

### 3. 5      結論

これらの検討から共通に言えることは、実験値との誤差は朗読調と会話調とで  
 は朗読調の制御規則を使っているため、大きな違いが現れた。誤差の最小値は、

朗読調で26.10%、会話調で38.16%であった。おおまかに言って、後者は前者の1.5倍の誤差が現れていることがわかった。

そしてまた、その誤差は、文末の強いニュアンスを持つ終助詞の母音部や、促音化している次の子音部に特に顕著に現れることがわかった。

#### IV 謝辞

この報告書を記すにあたって、樽松社長、嵯峨山室長、また、ご指導にあたってくださった匂坂さん、海木さん、その他多数のかたに感謝いたします。

#### IV 参考文献

[ 1 ] 阿部 匡伸、 匂坂 芳典、 桑原 尚夫: 「連続音声データベースにおける言語・韻律情報」、ATRテクニカルレポート、TR-I-0079

[ 2 ] 海木 延佳、 武田 一哉、 匂坂 芳典: 「文音声における音韻継続時間長の設定」、日本音響学会音声研究会資料

[ 3 ] 平戸 毅、 匂坂 芳典: 「種々の発話様式における韻律制御の検討」、ATRテクニカルレポート、TR-I-0149



音韻	朗読通常			会話通常			会話 朗読
	標本数	平均値	分散値	標本数	平均値	分散値	
a	144	90.3	19.2	129	83.6	39.1	92.6
i	70	73.6	18.1	68	55.6	22.3	75.5
u	44	50.9	11.4	37	45.1	24.7	88.6
e	86	87.3	15.8	75	75.7	39.6	86.7
o	74	84.3	17.4	65	82.3	34.3	97.6
s	35	104.9	29.7	31	86.3	18.8	82.3
sh	8	141.9	14.9	7	111.4	16.8	78.5
h	8	81.6	20.2	8	51.3	18.3	62.8
f	4	77.5	27.9	4	73.1	22.8	94.3
ts	10	119.2	19.1	9	81.9	22.8	68.7
ch	2	125.0	15.0	1	85.0	0	68.0
p	1	165.0	0	1	145.0	0	87.8
t	22	94.3	26.9	20	62.3	22.7	66.0
k	82	85.1	30.3	72	62.7	21.1	73.7
b	12	49.6	8.5	8	43.4	15.7	87.5
d	35	45.2	13.1	30	38.6	12.9	85.3
g	10	50.8	9.2	10	41.8	8.9	82.3
m	26	52.7	10.7	26	39.8	10.8	75.5
n	37	45.1	12.8	32	35.9	13.8	79.6
j	5	138.0	40.3	5	106.0	36.5	76.8
z	6	76.7	17.5	4	71.3	30.8	92.9
w	3	53.3	6.2	2	37.5	4.9	70.3
y	9	49.4	15.1	7	44.3	13.2	89.6
r	41	18.5	4.7	38	16.4	4.5	88.6

表1 朗読通常、会話通常の各音韻毎の統計値

発話 様式	予測誤差(ms)			誤差 / 平均 (%)			平均長 (ms)
	均一	要素毎	音韻毎	均一	要素毎	音韻毎	
朗読通常	24.50	22.67	21.12	32.18	29.79	27.76	76.08
遅い	36.56	34.78	33.84	35.48	33.76	32.85	103.02
速い	17.73	17.21	16.01	29.72	28.84	26.84	59.67
会話親切	44.17	43.89	42.16	49.88	49.58	47.62	88.53
怒り	27.64	26.13	24.94	47.70	46.02	43.93	56.77
急ぎ	24.73	24.10	22.68	46.01	44.95	42.30	53.61
通常	28.11	27.94	26.60	44.84	43.67	41.57	63.98

表2 品詞誤差補正なしの均一、要素毎、音韻毎の各変換における誤差

発話 様式	予測誤差(ms)			誤差 / 平均 (%)			平均長 (ms)
	均一	要素毎	音韻毎	均一	要素毎	音韻毎	
朗読通常	24.28	21.79	20.55	31.91	28.63	26.99	76.08
遅い	36.20	34.29	33.07	35.16	33.28	32.10	103.02
速い	17.60	15.91	15.59	29.49	26.66	26.10	59.67
会話親切	44.06	40.51	39.06	49.77	45.76	44.12	88.53
怒り	27.70	23.46	21.66	47.79	41.32	38.16	56.77
急ぎ	24.37	21.21	20.95	45.45	39.47	38.99	53.74
通常	28.05	26.59	26.30	43.83	41.66	41.21	63.83

表3 品詞誤差補正後の均一、要素毎、音韻毎の各変換における誤差

発話 様式	予測誤差(ms)			誤差 / 平均 (%)		
	1+3→1	1+3→3	1+3→2	1+3→1	1+3→3	1+3→2
朗読通常	19.71	24.42	25.93	27.06	31.73	32.98
遅い	28.64	38.91	39.28	29.03	36.99	37.54
速い	13.44	16.06	22.12	22.32	28.17	34.84
会話親切	37.24	42.54	40.60	40.44	48.81	46.88
怒り	23.29	27.95	24.88	41.69	48.43	44.23
急ぎ	18.35	23.23	26.88	35.16	42.35	49.85
通常	21.52	29.13	28.94	35.71	44.41	44.89

表4 全音韻均一の変換における誤差

発話 様式	予測誤差(ms)			誤差 / 平均 (%)		
	1+3→1	1+3→3	1+3→2	1+3→1	1+3→3	1+3→2
朗読通常	19.67	21.64	22.41	27.02	28.10	28.50
遅い	27.73	36.14	36.34	28.10	34.36	34.73
速い	15.40	15.24	20.27	22.26	26.72	31.94
会話親切	37.21	42.40	38.41	40.41	48.65	44.35
怒り	21.01	25.22	22.46	38.15	39.90	39.94
急ぎ	17.62	21.88	25.24	33.77	39.89	46.82
通常	20.49	28.41	27.96	34.01	43.27	42.69

表5 音韻種類毎の変換における誤差

発話 様式	予測誤差(ms)			誤差 / 平均 (%)		
	1+3→1	1+3→3	1+3→2	1+3→1	1+3→3	1+3→2
朗読通常	19.69	21.47	20.53	22.04	27.91	26.12
遅い	27.59	35.76	34.40	27.96	34.00	32.88
速い	13.37	14.82	19.54	22.22	25.99	30.79
会話親切	37.19	42.14	37.17	40.39	48.35	42.92
怒り	18.63	23.92	21.89	33.34	41.46	38.93
急ぎ	17.60	21.40	24.45	33.72	39.03	45.45
通常	20.41	27.87	27.49	33.87	42.44	42.62

表6 音韻毎の変換における誤差

	音韻数			音韻長		
	会話1	会話2	会話3	会話1	会話2	会話3
朗読通常	243	208	323	72.81	78.62	76.94
遅い	242	210	329	98.67	104.62	105.18
速い	242	203	341	60.19	63.47	57.03
会話親切	239	199	338	92.07	86.60	87.15
怒り	228	194	327	55.87	56.23	57.70
急ぎ	225	218	309	52.18	53.91	54.84
通常	207	178	334	60.25	64.49	65.66

表7 各会話データの音韻数、音韻長

	標本数	a	e	o
朗読通常	9	-74.87	-84.05	-34.26
遅い	8	-59.74	-80.87	-31.45
速い	9	-89.73	-86.43	-81.77
会話親切	9	-118.5	-136.8	-100.8
怒り	9	-10.73	-17.95	-10.33
急ぎ	9	-5.28	-17.82	-17.03
通常	9	1.95	-15.53	-11.67

表8 終助詞の誤差における影響

	標本数	平均誤差(ms)
朗読通常	9	-36.4
遅い	9	-152.2
速い	9	15.8
会話親切	9	107.8
怒り	9	70.9
急ぎ	9	70.6
通常	9	107.8

表9 促音化している部分の誤差

誤差 (ms)

-150			1
-145	1		
-140			
-135			
-130			
-125			1
-120	1		
-115			
-110	2		2
-105			1
-100			2
-95			2
-90	3	1	1
-85	2		
-80	1		1
-75			2
-70	1		2
-65	1		3
-60	1		1 3 1
-55	1		2 1
-50	1		1 8
-45	1		7
-40	2		3 3
-35	2		9 1
-30	7		1 9 1
-25	10		1 18 1
-20	23	1	2 21
-15	1 30	1	3 26
-10	3 34		4 40
-5	1 49	4	5 54 3
0	5 70	5	5 62 3
5	1 98	4	6 56 2
10	61		8 45
15	2 45	1	4 36 1
20	2 29	4	2 23
25	30		3 24
30	15		2 13 1
35	1 10	1	3 7 1
40	3 3		3 1
45	4		1 3
50	1 2		1
55	1		
60			
65			
70			
75			
80			
85			
90			
95			
100	1		
105			
110			
115			

a b c

a b c

会話調 急ぎ

会話調 通常

- a : 実測長が予測長より 30ms以上長い時の誤差分布
- b : 実測と予測の誤差が 30ms以下の時の誤差分布
- c : 予測長が実測長より 30ms以上長いときの誤差分布

図2 一つ前の音韻の誤差が30msを越えたときの次の音韻の誤差分布

