

TR-I-0195

言語情報を利用した母音継続時間長の制御

Vowel Duration Control
using Linguistic Information

海木 延佳

Nobuyoshi KAIKI

武田 一哉 匂坂 芳典

Kazuya TAKEDA Yoshinori SAGISAKA

1991.2.20

概要

自然な文音声を規則により合成するため、母音継続時間長の制御の精密化を図った。これまで日本語の母音継続時間長について指摘されてきた制御要因に加え、新たに、句構造や語彙等の言語情報が制御に及ぼす影響を調べるため、各母音に対して句内位置と形態素情報(品詞)を要因に加えて重回帰分析(数量化I類)を行った。分析には、4名が発声した503文章のべ2012文章中の母音セグメント約43000サンプルを用いた。この結果、従来主として単語音声の分析で指摘されてきた要因の文中における影響が定量的に明らかになった。また、新たに次の事柄が判明した。(1)従来知られている呼気段落よりも小さいアクセント句レベルで句全体の伸縮、句末の伸長が行われる。(2)数詞、固有名詞等の重要な情報を担う箇所が伸長される。(3)拍タイミング等による影響は、隣接音韻のみならず2音韻前後にも渡る。これら制御要因による推定精度を調べるため、全データを各人毎に2つのデータセットに分け、各々のセットに対して評価した。この結果、分析に利用した母音セットでは平均2乗誤差15.30ms(標準偏差の57.2%)、分析外母音セットに対して、平均2乗誤差15.84ms(標準偏差の59.2%)となり、規則の妥当性を確認できた。

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

©ATR Interpreting Telephony Research Laboratories

©ATR自動翻訳電話研究所

目 次

1. まえがき	2
2. 母音長制御要因の分析	3
2.1 分析用言語音声データ	3
2.2 制御要因	4
2.3 分析方法	5
2.4 分析結果	6
(1) 発話区分内モーラ数による影響	6
(2) 発話区分内位置による影響	6
(3) 品詞の種類による影響	6
(4) 近傍音韻の種類の影響	7
(5) アクセント核の有無の影響	7
3. 母音長推定実験	8
4. むすび	9
謝 辞	9
参考文献	10

1. まえがき

自然なリズムやテンポを持つ音声を規則によって合成するためには、音韻の継続時間長(以下、音韻長と呼ぶ)の制御が必要である。音韻長の制御には種々の要因が関与することが知られており、これまで提案されてきた制御規則化ではこれらの要因別に分析を行い、それらを統合する形式が取られてきた(1)-(3)。またこれらの分析は主として単語音声を用いて行われ、その値が文音声の制御にも用いられてきた。このような規則化が行われてきた背景には、制御要因が多岐にわたり、それらの分析にある程度十分な文音声データの構築が難しかったことがある。このような状況も災いして、日本語については、構文構造や意味情報等、文章レベルにおける音韻長の制御について分析が始められてから日が浅い(4)。

本論文では音韻バランスを考慮した言語音声データを用い、言語情報が音韻長制御に与える影響を分析する。同時にこれまで単語音声を中心に調べられてきた種々の制御要因の影響を文音声について定量化する。分析方法としては、重回帰分析の一形態であり、従来音韻長制御モデル(4)(5)として用いられた林の数量化I類(6)を用い、音韻長変動が大きくサンプル数の多い母音を分析対象とした。最後に、この分析結果の有効性を確認するために行った母音長の推定実験について述べる。

2. 母音長制御要因の分析

2.1 分析用言語音声データ

文音声データとしては、4名のアナウンサー(男性3名A、B、C、女性1名F)が読み上げた503文章、計2012文章を用いた(7)。これらの文章は、音韻単位の出現頻度のバランスを考慮し、新聞・雑誌等約1万文章から選択されたものであり、日本語に現れる可能性のある全ての2音韻連鎖と、鼻音化、無声化しやすいCVC連鎖、半母音を含むVCV連鎖を含んでいる(8)。このデータには、以下の情報を付与している。

- (1) 視察に基づく音韻境界、音韻の種類等の音韻ラベル(9)
- (2) 形態素境界、品詞、活用の種類、活用形
- (3) 文節境界、係受け関係
- (4) 聞き取りによるアクセント句境界、アクセント型

分析は、音韻長変動が大きくサンプル数の多い母音を対象とした。

2.2 制御要因

文章レベルの制御としては、これまで、呼気段落内モーラ数の増加に伴う短縮、呼気段落末における伸長、文末における短縮が知られている(3)(4)。本論文では、これらを組織的に調べるため、文、呼気段落、アクセント句、単語の4つの発話区分について、次の2つの制御変数を用いた。

(1) 発話区分内モーラ数

各母音の属す4つの区分それぞれの区分内モーラ数

(2) 発話区分内位置

各母音の属する区分内位置 (区分頭、区分中、区分末)

これら2つの制御変数は文を構成する単位のまとまりや単位境界での制御を考えたものであり、文の構造から決まるものであって、統語的制御要因といえる。

これに対し、語彙、意味情報による制御を反映する制御変数として、

(3) 品詞の種類

各母音の属す品詞(固有名詞・サ変名詞・形容名詞・普通名詞・数詞・代名詞・格助詞・準体助詞・係助詞・副助詞・並立助詞・接続助詞・終助詞・本動詞・補助動詞・形容詞・副詞・連体詞・接続詞・間投詞・感動詞・助動詞・接頭語・接尾語)の24種類を考えた。

これらに加え、従来、単語レベルの分析で明らかにされてきた制御変数として以下のものを用いた。

(4) 近傍音韻の種類

各音韻(/s/, /sh/, /ts/, /ch/, /p/, /t/, /k/, /h/, /f/, /m/, /n/, /j/, /z/, /w/, /y/, /b/, /d/, /g/, /r/, /a/, /i/, /u/, /e/, /o/, /N/, 拗音, ポーズ)

近傍は、サンプル数を考慮して2つ前後までの範囲(計4音韻)とした。

(5) 先行、後続子音の促音化

(6) アクセント核の有無

(7) 当該母音の種類(/a/, /i/, /u/, /e/, /o/)

これらのうち新たに、(4)近傍音韻については2つ前後の音韻の種類を、(5)先行、後続子音の促音化については先行子音の促音化を制御変数に加えた(4)。また、(6)アクセント核による影響は、従来、無意味単語を用いた測定(10)により単語レベルでは認められていないが、確認のため用いることにした。

2.3 分析方法

母音長制御モデルとしては、制御構造の階層性を反映させたもの(3)(11)(12)から全く計算式を仮定しない統計モデル(13)(14)まで種々提案されている。ここでは従来の知見との対比、統計的な扱いの必要性、計算の容易さの観点から、重回帰分析の一手法である数量化I類を用いる。母音長を次式によって表わすことを考える(4)(5)。

$$\hat{y}_i = \bar{y} + \sum_{f,c} x_{fc} \delta_{fc}(i), \quad (i = 1, \dots, N) \quad (1)$$

- ここで、 \hat{y}_i : i 番目の推定母音長、
 \bar{y} : 全データの平均母音長、
 N : 全データ数(約10800)、
 $\delta_{fc}(i) = 1$: i 番目のデータがファクタ f のカテゴリ c に属する時
 0 : それ以外の時、(2)

x_{fc} は $\sum_i (\hat{y}_i - y_i)^2$ を最小化することによって得られる(ただし、 y_i は i 番目のデータの

母音長)。ファクタ f には、文内モーラ数、呼気段落内モーラ数、文内位置、品詞の種類といった、2.2で述べた制御変数が対応し、カテゴリ c としては、例えばファクタ f が当該母音の場合、母音の種類(/a/, /i/, /u/, /e/, /o/)が対応する。

この分析で得られる x_{fc} は、カテゴリ c のスコアと呼ばれ、スコアの範囲がより大きいファクタは、母音長により大きな影響を与え、負(正)のスコアを示すカテゴリは、母音長において短縮(伸長)の影響を示す。分析で得られた平均母音長 \bar{y} 、各スコア x_{fc} を用い式(1)により、各ファクタ(制御変数)の線形加算によって推定母音長は算出される。

2.4 分析結果

分析は4名の話者各々に対して独立に行った。各制御変数の偏相関係数と重相関係数を各話者毎に表1に示す。この表より、先行音韻、次に後続音韻、当該音韻の順に影響が大きいことがわかる。

以下、各制御変数ごとに結果をまとめる。

(1) 発話区分内モーラ数による影響

発話区分内モーラ数による影響は、文を構成する区分を意識した制御の存在を意味する。話者によらず共通した傾向として、呼気段落内モーラ数の増加に対して母音長は減少する。また伸縮幅は小さいが、文、アクセント句内モーラ数の増加に対しても母音長減少の傾向がみられる、しかしながら単語にはこの傾向がみられない。これらは従来から示されている文、呼気段落に加えて、アクセント句程度の単位の制御が存在することを示している。

一例として話者Bについて、4つの発話区分(文、呼気段落、アクセント句、単語)内モーラ数による母音長の伸縮の値を図1(1)~(4)に示す。図の横軸はカテゴリ(モーラ数によって分けたグループ)を示し、縦軸は、平均母音長 \bar{y} (全母音の平均継続長)に対する、各カテゴリのスコア x_{fc} (伸縮の影響)を表している。

(2) 発話区分内位置による影響

発話区分内位置による影響は、文を構成する区分の境界による制御を示している。話者によらず共通した傾向として、文末の母音長は短く、呼気段落末は長くなることを確認した。また新たに、アクセント句末で伸長すること、単語区分内位置は影響がほとんどないことが確認できた。これらは、発話区分内モーラ数による傾向と符合し、アクセント句程度の区分による制御の存在が明らかとなった。4つの発話区分(文、呼気段落、アクセント句、単語)内位置による影響を図2(1)~(4)に示す。

さらに、発話区分境界による制御の影響の範囲について調べるため、発話区分末から3モーラの伸縮傾向を調べた。この結果、制御は発話区分末のモーラだけに限定されることが判明した。一例として図3に話者Bについての分析結果を示す。

(3) 品詞の種類による影響

図4に品詞の種類による母音長の伸縮傾向を示す。この図に示されるように、品詞の種類による伸縮幅は $-8\text{ms} \sim +12\text{ms}$ と小さいが、品詞の違いによる伸縮には統一した次のような傾向が観察された。

- (1) 英語における伸縮傾向(15)同様、全般的な傾向として、自立語(内容語)は付属語(機能語)に比べ伸長する。

- (2) 自立語のうちでも、特に情報を多く担っていると考えられる数詞や固有名詞等では、伸長傾向が顕著である。
- (3) 一方、付属語のうちでも、並列句等の句構造を示す上で重要な並列助詞や主題を提示するためよく用いられる「は」等を多く含む副助詞では、逆に伸長する。

日本語の場合、もともとモーラタイミングに起因する制約がきつく、英語ほど語彙に起因する伸縮は顕著ではないが、これらの伸縮傾向は言語に共通した制御要因の存在を示唆している。

(4) 近傍音韻の種類の影響

先に示した表1の偏相関係数が表すように、2つ前後の音韻による当該音韻への影響は無視できないことがわかる。このことより、時間長補償は単に隣接音韻に留まらず、その前後の音韻(2つ前後の音韻)にまで及ぶことが推察される。

(5) アクセント核の有無の影響

偏相関係数を示した表1により、4名の話者において、アクセント核の有無による要因の偏相関係数は、0.006から0.018と他の制御変数に比べて極めて低い値となっている。このことより、アクセント核の有無は、制御にほとんど関与していないといえる。この結果は、種々の発話様式における韻律制御において、時間長制御が基本周波数と振幅の制御とは独立して行われるという知見(16)と一致する。

3. 母音長推定実験

前章の検討結果をもとに、母音長制御変数を以下のように選んだ。

- (1) 発話区分内モーラ数 (文、呼気段落、句)
- (2) 発話区分内位置 (文、呼気段落、句)
- (3) 品詞の種類
- (4) 近傍音韻の種類 (前後、2つ前後)
- (5) 先行、後続子音の促音化
- (6) 当該母音の種類

4名の話者それぞれについて503文を2つのデータセットに分け、各々のセットについて分析し、各制御変数 f のスコア x_{fc} を求める。次に、これらの値を用いて分析に利用したデータセット(学習データ)と分析に利用しなかったデータセット(未知データ)の母音長を式(1)により推定し、それらの測定値との誤差を求めた。

表2に2つのセット毎の平均2乗誤差を示す。4名の話者の平均で、学習データに対して、平均2乗誤差 15.30ms(平均母音長に対して19.6%、標準偏差の57.2%)、また未知データに対して、平均2乗誤差 15.84ms(平均母音長に対して19.9%、標準偏差の59.2%)の結果を得た。この結果が示すように未知データに対する推定誤差は学習データのものに近く、ほぼ同様の精度で推定できることが判明した。

次に、大きな予測誤差を生ずる原因を明らかにするため、話者Aについて測定値と推定値の差が30msを越える498サンプル(全体の4.5%)を調べた結果、次のようなものであることが判明した。((1)~(4)は重複するサンプルを含む)

- (1) 母音-母音(撥音を含む)の連鎖。(241サンプル)
- (2) 発話区分末。(文末21サンプル、呼気段落末87サンプル、アクセント句末200サンプル)
- (3) 呼気段落頭。(35サンプル)
- (4) 前後の音節が促音。(44サンプル)
- (5) その他。(103サンプル)

これら大きな推定誤差を生ずるものとしては、音韻境界の測定誤差に起因すると思われるものも少なくない(特に(1)、(2)中)が、誤差傾向の偏りは次の制御の必要性を示唆している。

- (a) 音節を単位とした時間長制御
(1)にみられるような /ai/、/aN/ 等、2モーラ1音節となるいわゆる2重母音と単なる母音連鎖の区別。(4)の促音節の取扱い。
- (b) 発話区分境界
(2)、(3)にみられるように単位境界、特に区分末にみられる単語依存性。

4. むすび

本論文では、自然なリズムやテンポを持つ日本語音声規則によって合成するために、構文構造や意味情報等が音韻長制御に与える影響の定量分析を行った。その結果、以下の制御の存在が定量的に明らかになった。

- (1) アクセント句程度の発話区分による制御が存在する。このため、アクセント句末が長くなり、アクセント句内のモーラ数増加に対し母音長は減少する。
- (2) 英語同様、日本語でも重要な情報を担う箇所は伸長する。しかしこの伸縮の割合は英語に比べて小さい。
- (3) モーラタイミング等による音韻継続時間長補償は単に隣接音韻に留まらず、その前後の音韻(2つ前後の音韻)にまで及ぶ。
- (4) アクセント核の有無は母音長に影響をほとんど与えない。

さらに、これら分析結果の有効性を確認するため、これらの要因を用いた母音長推定実験を行った。推定実験の結果、学習データに対して平均2乗誤差15.30ms(平均母音長の19.6%)、未知データに対して平均2乗誤差15.84ms(平均母音長の19.9%)の精度で推定を行えることを示した。

今後に残された課題として、(1)音節構造を考慮した母音接続部、特殊拍の時間長制御、(2)子音、無声化母音等への拡張、(3)制御モデルの改良、(4)発声話者による相違の分析、などを行っていくことが必要と思われる。

謝 辞

研究の機会を与えて下さった樽松社長、嵯峨山室長、NTTヒューマンインタフェース研究所の鹿野博士に感謝いたします。また、討論していただいた音声情報処理研究室の諸氏に感謝します。

参考文献

- (1) Klatt, D. H. : "Linguistic uses of segmental duration in English : Acoustic and perceptual evidence", J. Acoust. Soc. Am., 59,5, pp.1208-1221 (1976).
- (2) 樋口宣男、藤崎博也：“連続音声における音素的特徴の時間制御”、音響学会音声研資、S80-40, (1980).
- (3) 匂坂芳典、東倉洋一：“規則による音声合成のための音韻時間長制御”、信学論(A)、J67-A, 7, pp.629-638 (昭59-07).
- (4) Takeda K., Sagisaka Y. and Kuwabara H. : "On sentence-level factors governing segmental duration in Japanese", J. Acoust. Soc. Am., 86,6, pp.2081-2087 (1989).
- (5) 酒寄哲也、佐々木昭一、北川博雄：“規則合成のための数量化1類を用いた韻律制御”、音響学会講演論文集、3-4-17 (昭61-10).
- (6) 駒澤勉：“数量化理論とデータ処理”、朝倉書店、(1982).
- (7) 阿部匡伸、匂坂芳典、桑原尚夫：“言語・韻律情報を持つ連続音声の基本周波数データベース”、音響学会講演論文集、2-3-22 (平1-10).
- (8) 磯健一、渡辺隆夫、桑原尚夫：“音声データベース用文セットの設計”、音響学会講演論文集、2-2-19 (昭63-3).
- (9) Kurematsu A., Takeda K., Sagisaka Y., Katagiri S., Kuwabara H. and Shikano K. : "ATR Japanese Speech Database as a Tool of Speech Recognition and Synthesis", Speech Communication Vol.9, No.4, pp357-364 (1990).
- (10) 宮原末治、橋本新一郎：“単語の時間構造に関する一考察”、信学全大、1581, pp1432 (昭49).
- (11) 佐藤大和、竹内晟吉：“音声合成のための音韻継続時間の規則について”、音響学会講演論文集、3-5-7 (昭53-10).
- (12) Campbell W. N. : "Analog I/O Nets for Syllable Timing", Speech Communication., Vol.9, pp.57-61 (1990).
- (13) Riley M. D. : "Tree-based Modelling for Speech Synthesis", ESCA Speech Synthesis Workshop (1990-9).
- (14) Pitrelli J. F. : "Hierarchical Modeling of Phoneme Duration: Application to Speech Recognition", Ph. D. Thesis, Massachusetts Institute of Technology, (1989-5).
- (15) Umeda N. : "Vowel duration in American English", J. Acoust. Soc. Am., 58, pp.434-445 (1975).
- (16) 宮武正典、匂坂芳典：“種々の発話様式に見られる韻律特徴とその制御”、信学論(D-II)、J73-D-II, 12, pp.1929-1935 (1990-12).

表1 各話者、各制御変数の偏相関係数

制御変数	偏相関係数				
	話者	A	B	C	F
当該母音		0.475	0.470	0.364	0.450
近傍 前後 2つ前 2つ後	前後	0.564	0.592	0.488	0.592
	前後	0.483	0.409	0.416	0.483
	2つ前	0.210	0.336	0.167	0.221
	2つ後	0.233	0.225	0.197	0.176
促音 前後	前後	0.007	0.045	0.020	0.042
	前後	0.130	0.036	0.132	0.035
位置 文 呼気段落 句 単語	文	0.428	0.524	0.323	0.282
	呼気段落	0.218	0.406	0.486	0.121
	句	0.170	0.114	0.157	0.075
	単語	0.012	0.013	0.022	0.062
モーラ 数 文 呼気段落 句 単語	文	0.069	0.083	0.068	0.055
	呼気段落	0.146	0.119	0.139	0.143
	句	0.085	0.091	0.085	0.055
	単語	0.042	0.036	0.060	0.073
アクセント核の有無		0.015	0.003	0.018	0.006
品詞		0.115	0.166	0.171	0.133
重相関係数		0.798	0.863	0.796	0.769

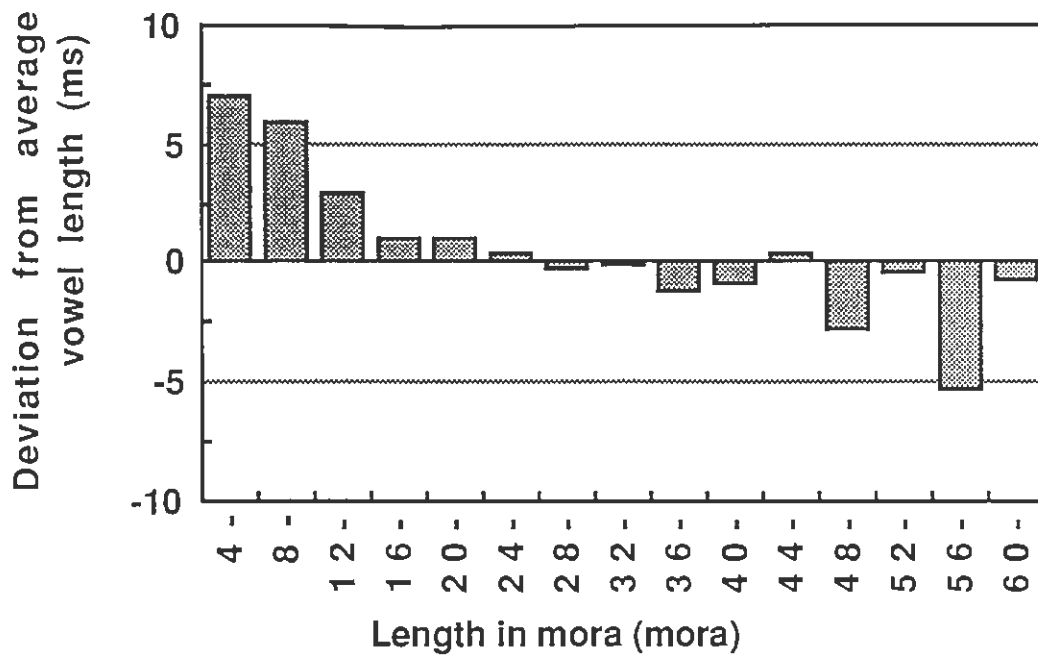


Figure 1(1). Decrease of vowel duration in inverse proportion to the increase of mora count of sentence (Speaker B)

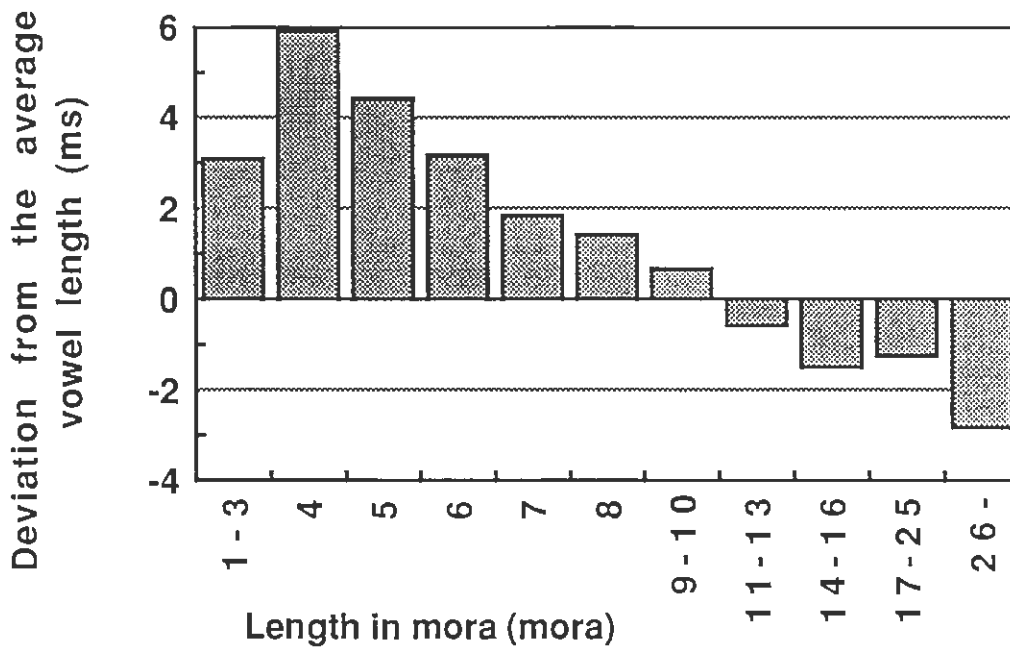


Figure 1(2). Decrease of vowel duration in inverse proportion to the increase of mora count of breath group (Speaker B)

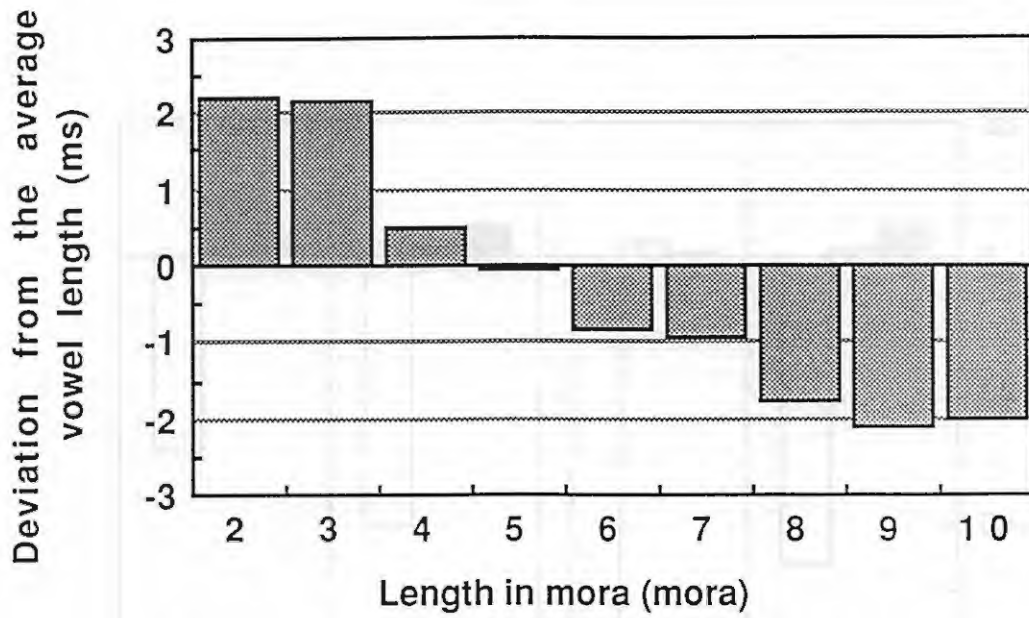


Figure 1(3). Decrease of vowel duration in inverse proportion to the increase of mora count of phrase (Speaker B)

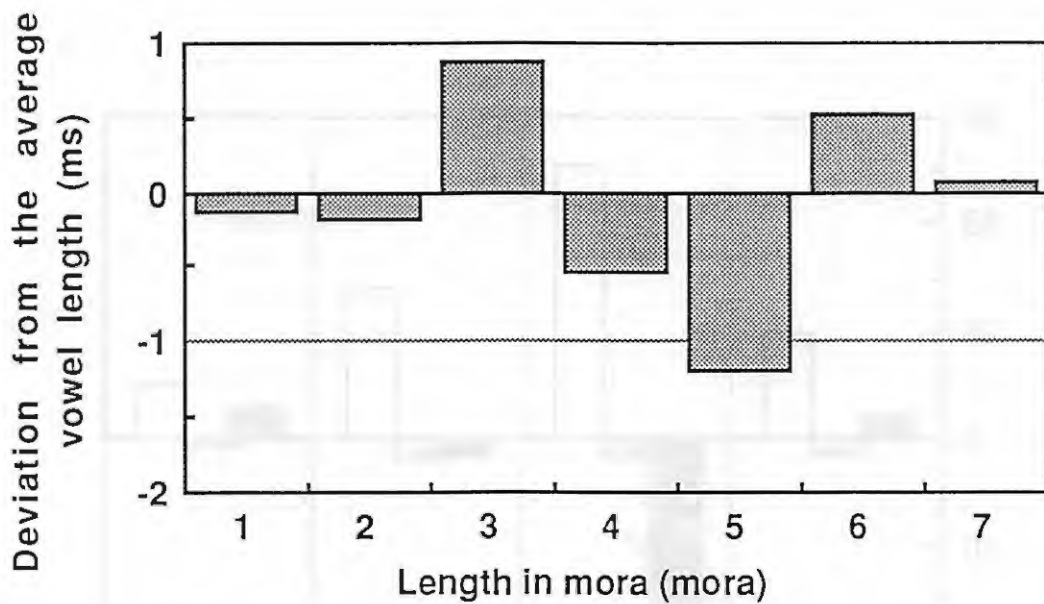


Figure 1(4). No effect of vowel duration of mora count of word (Speaker B)

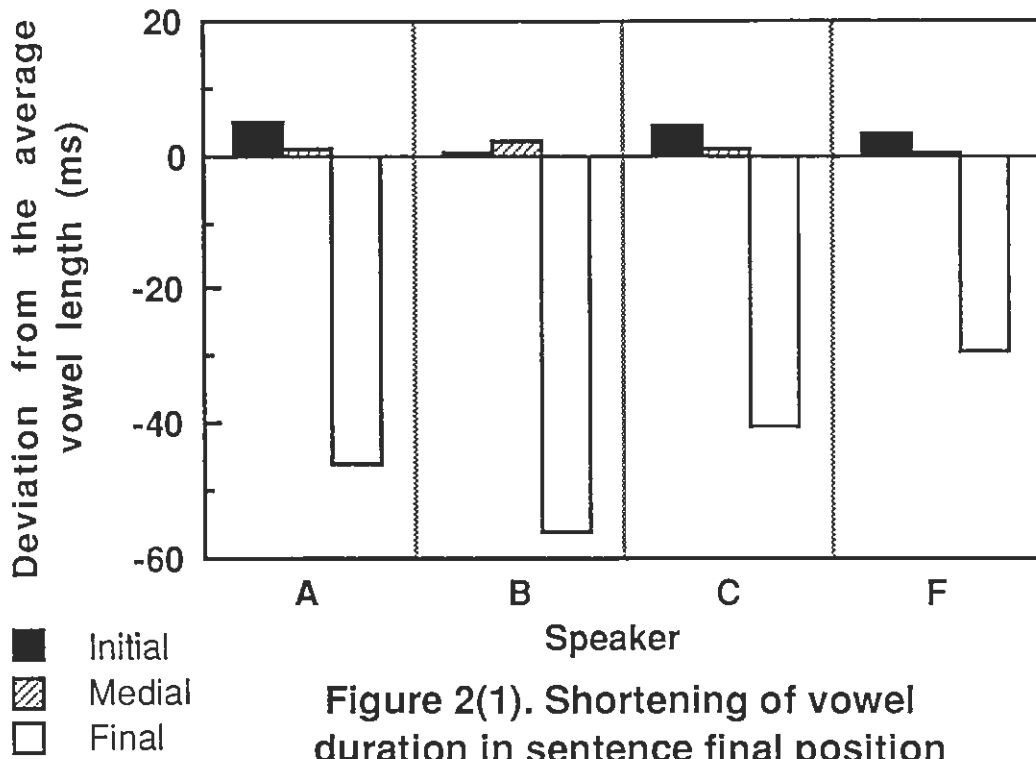


Figure 2(1). Shortening of vowel duration in sentence final position

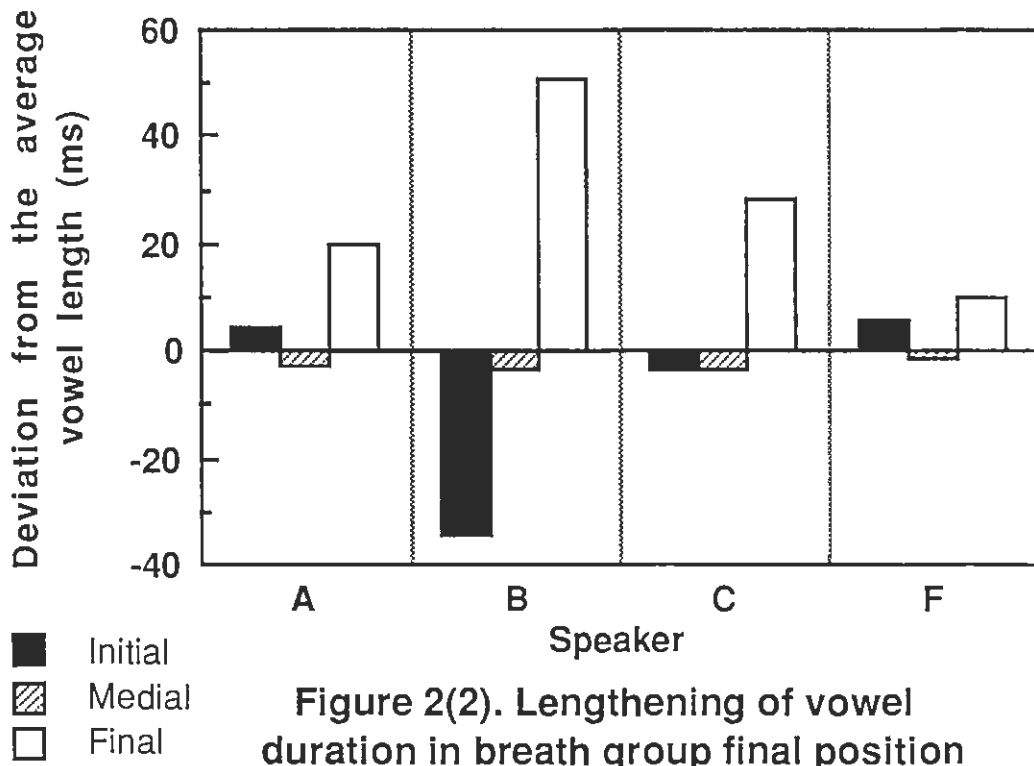


Figure 2(2). Lengthening of vowel duration in breath group final position

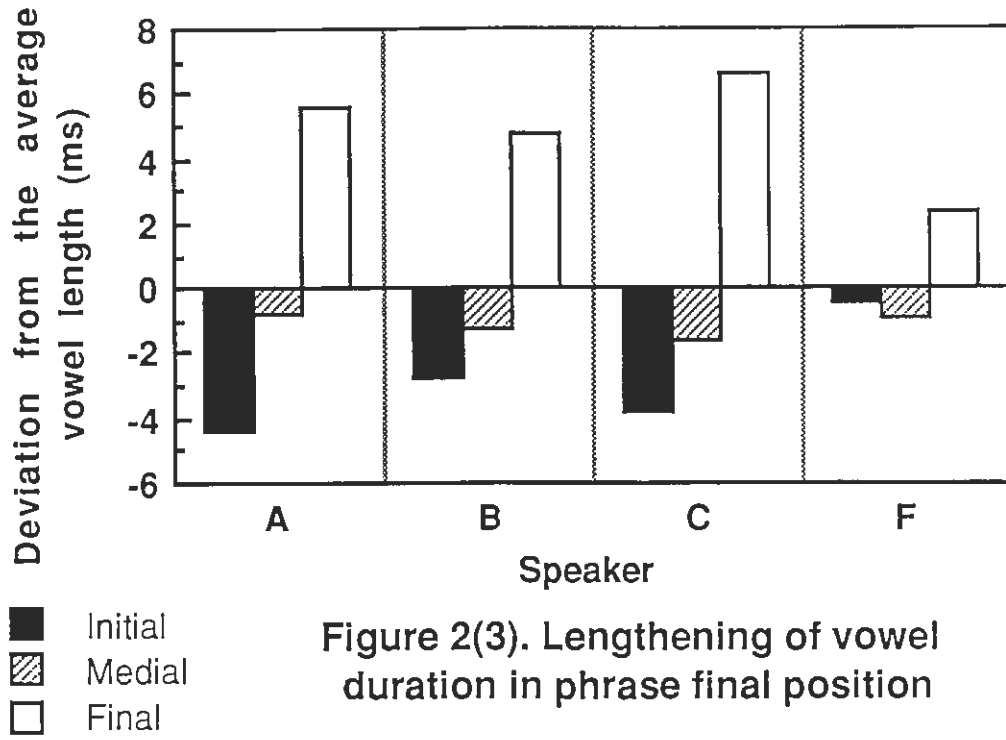


Figure 2(3). Lengthening of vowel duration in phrase final position

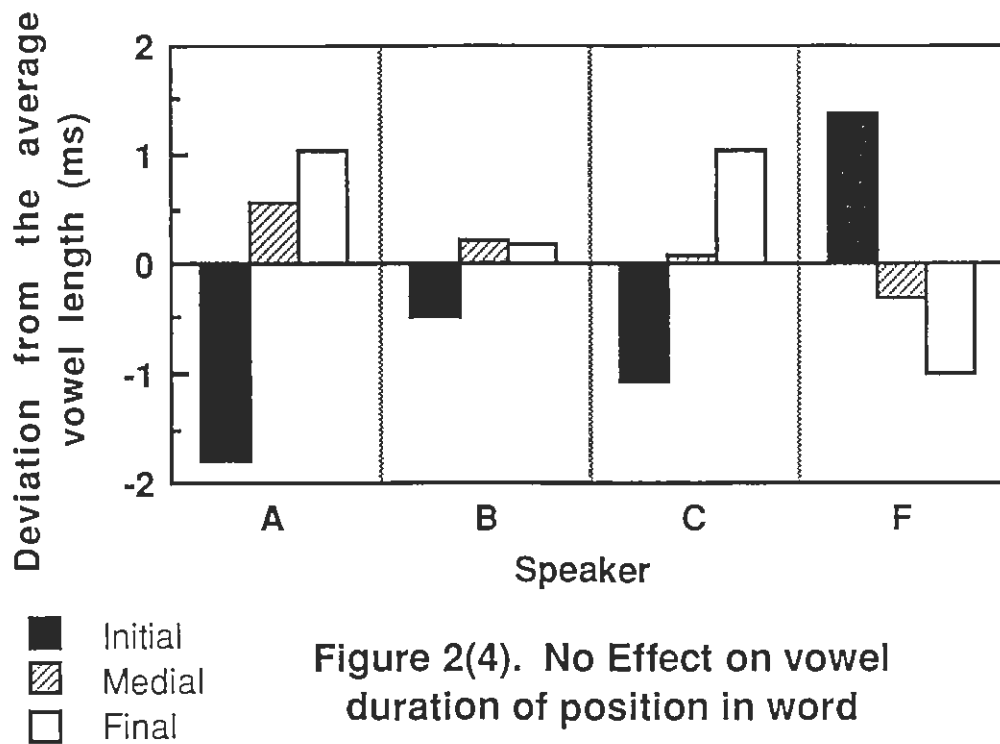


Figure 2(4). No Effect on vowel duration of position in word

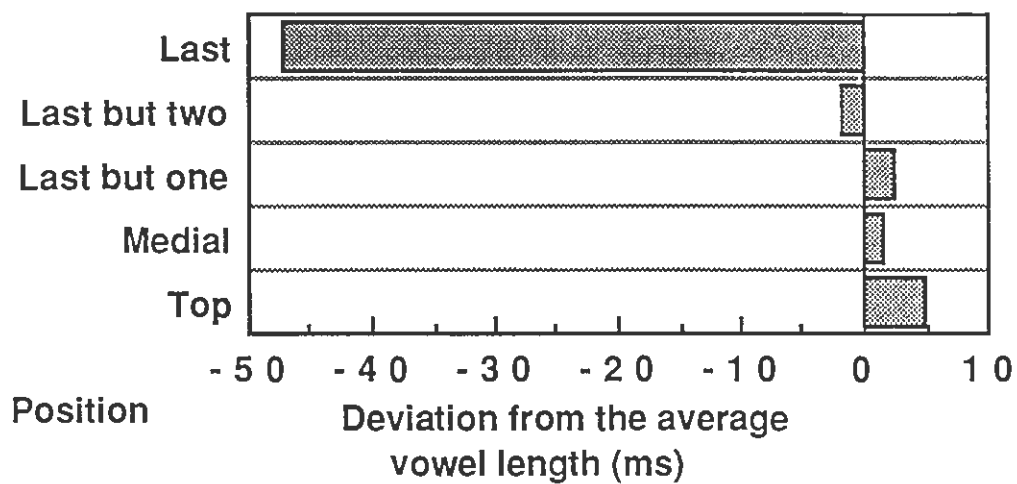


Figure 3(1). Shortening of only the final mora in sentence (Speaker B)

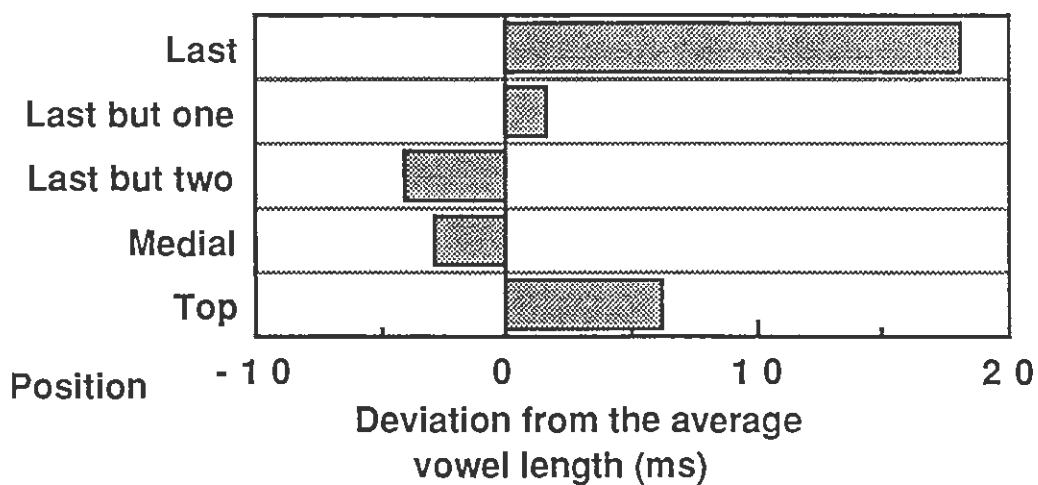
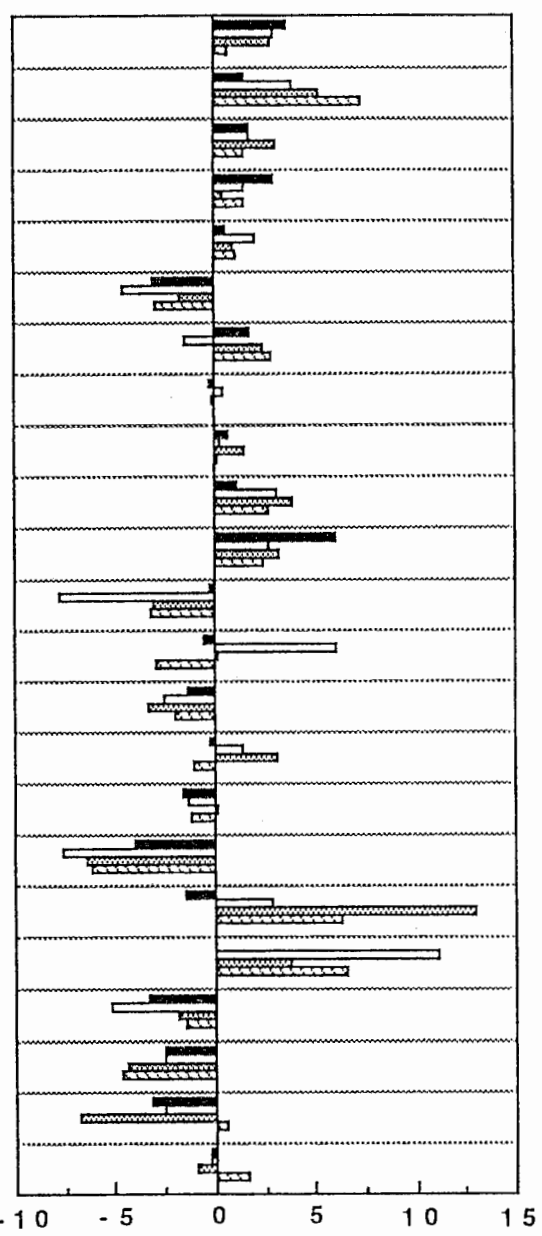


Figure 3(2). Lengthening of only the final mora in breath group (Speaker B)

proper noun	(固有名詞)
quantifier	(数詞)
nominal noun	(普通名詞)
adjectival noun	(形容名詞)
verbal noun	(サ変名詞)
pronoun	(代名詞)
prefix	(接頭語)
suffix	(接尾語)
adverb	(副詞)
adjective	(形容詞)
coordinate particle	(並列助詞)
final particle	(終助詞)
modal particle	(係助詞)
case particle	(格助詞)
adverbial particle	(副助詞)
adnominal particle	(準体助詞)
conjunctive particle	(接続助詞)
conjunction	(接続詞)
exclamation	(感動詞)
interjection	(連体詞)
auxiliary verb	(助動詞)
attributive	(補助動詞)
verb	(本動詞)



Part of speech

Deviation from the average vowel length (ms)

Speaker
 ■ A
 □ B
 ▨ C
 ▩ F

Figure 4. The differences of vowel duration between content word and function words

表2 母音長推定評価結果
 測定値と推定値の平均2乗誤差 [ms],
 (平均母音長に対する割合 [%])
 (標準偏差に対する割合 [%])

データ	話者A	話者B	話者C	話者F	平均
学習 データ	14.58 (19.2) (60.0)	14.90 (18.1) (50.0)	16.38 (22.7) (55.7)	15.32 (18.4) (63.0)	15.30 (19.6) (57.2)
未知 データ	14.87 (19.6) (61.2)	15.47 (18.8) (51.9)	17.08 (23.7) (58.1)	15.92 (19.2) (65.5)	15.84 (19.9) (59.2)