TR-IT-0123

回帰木を用いる韓国語音声合成の音韻継続時間のモデル化

Segmental duration modeling for Korean text-to-speech synthesis using regression tree

李亮熙

Yang Hee LEE

1995.08

音声合成において、自然なリズムの音声を合成する為には、音韻継続時間制御は必要であり、特に、音節の長さによって意味を弁別する韓国語においては最も重要である。従って、韓国語の時間的な特徴を分析し、タイミング制御のため音韻継続時間をモデル化した。韓国語音声合成における音韻継続時間の制御規則を生成するために、韓国語の次のような時間的特徴に対して分析を行なった。合成音に発話テンポに関係なく自然なリズムを与えるために、発話テンポの変化による音韻とポーズの継続時間の変化を調べた。また、音韻継続時間において、音節タイプによる母音長の変化、文節と句における音節の位置と数、 そして隣接する音節タイプと音韻の影響に対して統計的に分析した。さらに、これら特徴を制御ファクタとして用いた音韻継続時間予測モデルを生成し、このモデルで音韻継続時間の予測実験を行なった。

目次

表目次

図目次

1 はじめに

自然なリズムとテンポを持つ音声を合成するために、精巧な音韻継続時間のモデル化が必要である。多くの言語に対して、音韻継続時間のモデル化の研究が行なわれている [1],[3],[7],[6]。音声データから継続時間制御規則を自動的に生成するために、最近、統計的な方法が用いられている。回帰木モデル [6]、線形回帰モデル [3] とサム - オブ - プロダクト モデル [6] などの方法が規則の最適化または、規則の自動生成のために提案されている。しかし、音節の長短によって意味を弁別する韓国語において、韻律の他の要素よりも継続時間制御モデルは特に重要であるが、継続時間モデリングの研究があまり行なわれていない。最近、韓国語としては始めて、母音における前後隣接する両方の 2 つ子音の効果を調べ、Klatt モデルに適用したが、各々音素の影響を正確に調べなかった。従って、ここでは正確な音韻継続時間のモデル化を図るため、韓国語のタイミング特徴に対して統計的に正確に分析を行なう。どのような発話テンポでも自然な音声を生成するために、1) 発話テンポによるポーズ継続時間と音韻の継続時間の変化を調べる。2) 音韻環境に対して、前後音節タイプによる音節継続時間の変化及び前後音韻による影響を調べる。3) 文節または句内の音節位置による効果を調べる。これらのタイミング特徴を用いて統計的方法である回帰木で音韻継続時間をモデル化する。

2　韓国語の音節構造と音素

韓国語の文字、ハングルは、表音文字で3つの要素（初声：子音、中声：母音、終声：子音）の組合せで構成される。つまり、次の式のように表す。

$$(C_i) + V + (C_f), \tag{1}$$

C_i　：　初声（文字素：19、音素：18）
V　：　音節核（単母音：9、半母音（w、y）＋単母音：11、複母音：1）
C_f　：　終声（文字素：27、音素：7）

ここで、（）は、（）内の文字素（音素）がなくても、文字として成立する場合もあることを表している。また、初声には、V と VC 型の文字（音節）を表すための冗長な文字素を含めて、19種類がある。さらに、終声（子音）においては27種類の文字素があるが、これは音韻としては、7種類の音素に帰着する。
このように、音素が音韻環境により他の音素に変わる音韻変動や終声の文字素と音素間の上へのマッピングである音韻帰着が起こるため、文字表記と音韻表記と異なる。従って、テキストから音声を合成する場合、正書法文表記から音素表記への変換が必要である。次は音素表記をラベリング記号で表した例である。

baraMgwa hANnimi sEro himi sedago datugo iSIR TA, haN nagInega TaTItaN weturIR iBKo gErEwaSIMnida.
（風とお日さまがお互いに力が強いと争っている時、一人の旅人が暖かい外套を着て歩いて来ました.）

韓国語の音素と各々に対応するラベリングのための記号、各音素の調音様式と調音位置を表2に示す。

また、これらの音素の中で、ある音素は発声環境によって音素自体は同じであるが音韻性が変わる音韻の変異が起こる場合が多い。その中で、代表的な初声の音韻変異（allophone）は次のようである。

/g,b,d/　：　前に pause ある場合 - voiceless unaspired lenis plosive
/g,b,d/　：　前に有声音 - voiced unaspired lenis plosive
/z/　：　前に pause ある場合 -voiceless unaspired lenis affricative
/z/　：　前に有声音 - voiced unaspired lenis affricative
/h/　：　前に pause ある場合 -voiceless fricative
/h/　：　前に有声音 - voiced fricative

表 1: Korean phoneme table

音節要素	音素 ラベリング 対応記号	IPA 記号	調音様式	調音位置	その他
初声 (子音)	g	g	plosive	velars	lenis
	n	n	nasal	alveolars	
	d	d	plosive	alveolars	lenis
	r	l	liquid	alveolars	
	m	m	nasal	labials	
	b	b	plosive	labials	lenis
	s	s	fricative	alveolars	lenis
	z	z	affricate	palatals	lenis
	c	c^h	affricate	palatals	aspiration
	k	k^h	plosive	velars	aspiration
	t	t^h	plosive	alveolars	aspiration
	p	p^h	plosive	labials	aspiration
	h	h	fricative	glottal	
	K	k	plosive	velars	fortis
	T	t	plosive	alveolars	fortis
	P	p	plosive	labials	fortis
	S	s	fricative	alveolars	fortis
	Z	c	affricate	palatals	fortis
終声 (子音)	G	k	plosive	velars	coda
	N	n	nasal	alveolars	coda
	D	t	plosive	alveolars	coda
	R	l	liquid	alveolars	coda
	M	m	nasal	labials	coda
	B	p	plosive	labials	coda
	Q		nasal	velars	coda

表 2: Korean phoneme table continue

中声	a	a	low	central	unrnd
(音節核)	E		mid	central	unrnd
	o	o	higher mid	back	round
	u	u	high	back	round
	I		high	central	unrnd
	i	i	high	front	unrnd
	A		lower mid	front	unrnd
	e	e	higher mid	front	unrnd
	O(w+e)	ϕ	higher mid	front	round
	y+a	ja	semivowel+vowel		
	y+E	j	semivowel+vowel		
	y+o	jo	semivowel+vowel		
	y+u	ju	semivowel+vowel		
	y+e	je	semivowel+vowel		
	y+A	j	semivowel+vowel		
	w+a	wa	semivowel+vowel		
	w+E	w	semivowel+vowel		
	w+A	w	semivowel+vowel		
	w+e	we	semivowel+vowel		
	w+i	wi	semivowel+vowel		
	U	i	compound vowel		

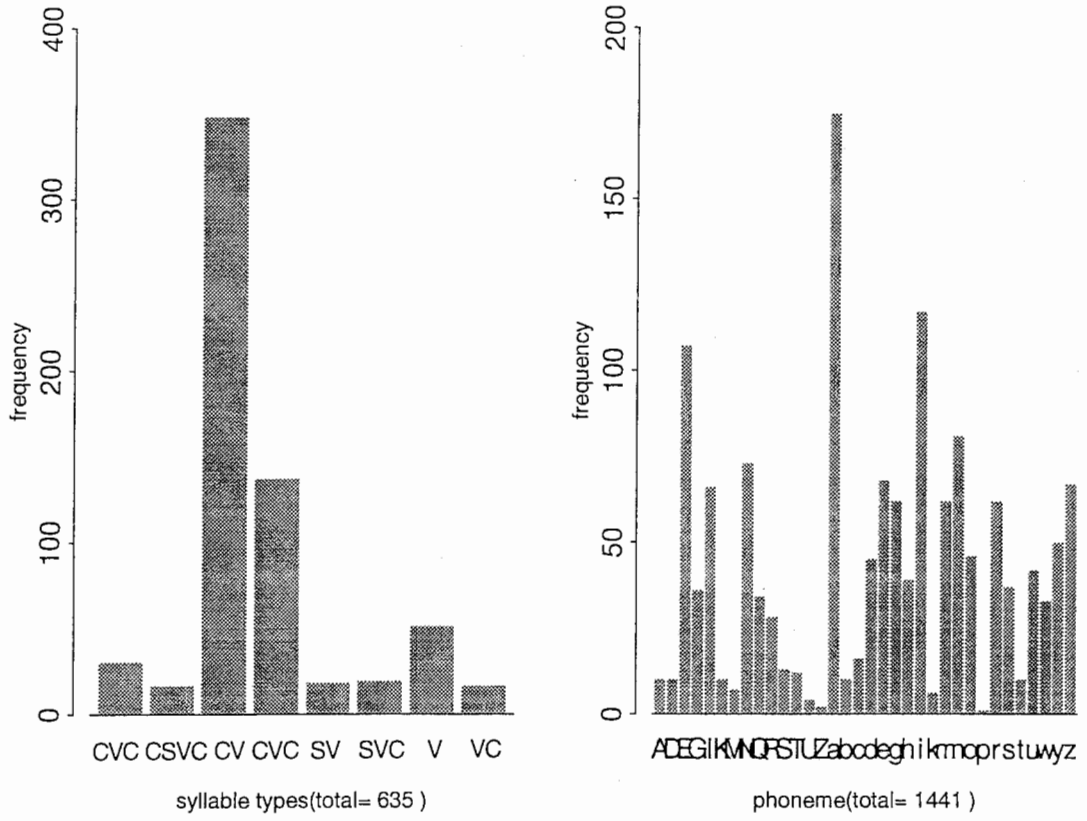

図 1: Distribution of syllable types and phoneme in speech data

3 韓国語のタイミング特徴分析

3.1 音声データ

タイミング特徴を分析するために用いられた音声データは、7 人の話者（韓国生まれの大学生、男性：3 人、女性：4 人）が、16 文章に対して、3 種類のテンポ（fast,normal, slow）で、読み上げた 336 発話である。この音声データに含まれている音節の数は 635 個であり、音素の数は 1441 個である。音節種類別の分布は図 1のように CV 型の音節が一番多く、次に CVC 型の音節である。図 1の音節タイプの表記において、S は半母音を含む音節を表す。また、音声データ内に含まれている音素の分布を図 2に表す。

3.2 発話テンポに対する分析

音声合成において、より高度な韻律制御モデル構築のため、まず、発話のテンポが韻律にどのように影響を及ぼすのかを調べる。テンポの変化に伴って、変化する特徴量と変化しない特徴量とをしることによってより良いモデルの実現が可能であると考えるからである。実際、テンポを変えても、韓国語としての自然なリズムが保たれている。このことから、テンポが変化しても時間構造を保つ単位が存在すると考えられる。テンポの変化により、変化が大きい要

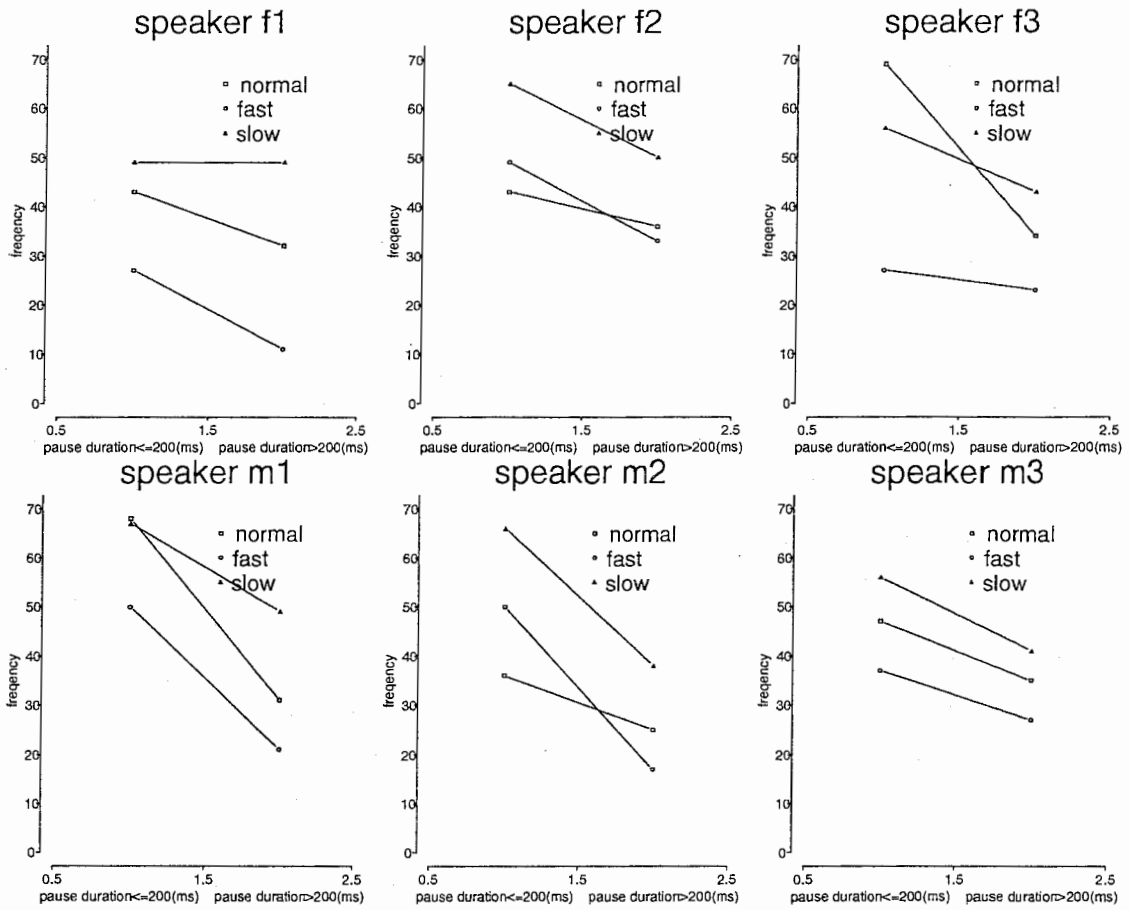

図 2: Distribution of pause duration

素と変化が小さい要素に分けて考えられる。ここでは、時間構造を音韻とポーズに分けて考える。まず、各々の発話のテンポの変化に対して、ポーズの継続時間長の分布を調べた。その結果、図2 ように、7 人の話者共 200ms 以上の長いポーズの数は、 低速＞中速＞高速の順であるが、短いポーズの数は必ずしも低速＞中速＞高速のような順ではない。このことは短いポーズより 200ms 以上の長いポーズの増減が発話のテンポの変化に影響を与えることを意味する。

また、発話のテンポの変化に同じ文を中速で読み上げた場合を基準として、高速及び低速で読み上げた場合の対応する音韻区間とポーズの継続時間長の伸縮率を測定した。即ち、次の式を用いて、各文に対して、音韻とポーズの継続時間長の伸縮を求めた。

$$phonerate = (SENdur_t) \div (SENdur_n), \tag{2}$$

$$pauserate = (PAUdur_t) \div (PAUdur_n), \tag{3}$$

ここで、$SENdur_t$ と $PAUdur_t$ は fast, slow テンポの発話（文）において各々ポーズを除いた音韻継続時間とポーズ継続時間を表し、$SENdur_n$ と $PAUdur_n$ は normal テンポの音韻継続時間とポーズ継続時間を表す。

話者7名の16文の伸縮率と各話者の平均伸縮率をもとめた。その結果から、高速の場合音韻の縮む率は話者によって多少異なるが大体10で、低速の場合、11の場合60ら、音韻よりポーズの方が発話のテンポの変化に大きく影響を受けることが分かる。各々の音韻は調音するのに最小限の固有の長さが保たなければならないから、縮む率が伸び率より小さいと考えられる。ポーズの縮む率でも最小限の呼気段落が必要であるので縮む率が伸び率より小さいと考えられる。

文内の音韻全体の伸縮率は大体10により音韻の弾性が異なるため、各音韻を分類して、各音韻グループの伸縮率を求め、合成の時、予測値に音韻グループの伸縮率を適用する。また、文構造により文節の接続度が求められ、それに従ってポーズの継続時間率を適用することにより自然な発話テンポを調節することが可能である。

3.3 音節継続時間の特徴

音節継続時間の特徴を分析するために、まず、シラブル-タイムド-リズムである韓国語において、各音節タイプごとに固有の音節継続時間を持つと考えられるので各音節タイプを構成する要素の継続時間、即ち母音と子音の継続時間を分析する。また、各音節継続時間に対する隣の音節による影響と文節及び句内の音節の数と位置による影響を調べる。これらの要素の影響を調べるために、音節継続時間に対する各音素の固有の継続時間の影響を排除するため、各音節タイプ別に各音素に対してZ Scoreで正規化を行なう。正規化音節継続時間は次の式で求める。

$$z_{ip} = (x_{ip} - \bar{x}_p) \div \sigma, \tag{4}$$

x_{ip} : 音素 p に対してi番目に観測された値
$\bar{x}_p$: 音素 p に対する平均値
σ : 音素 p に対する標準偏差

(3.3.1) 音節タイプにおける母音と子音の継続時間

韓国語音声のリズムに関して調べた文献によれば、若者においてシラブル-タイムド-リズムであると報告されている[2]。このことは各々の音節タイプごとに固有の音節継続時間を持つと言える。従って、各々の音節タイプを構成する各々の子音と母音は音節タイプによって固有の継続時間を持つのではないかと思われる。そのような関係を調べた結果を図3と図4に示す。これらの図においてSを含む音節タイプは半母音を含んでいる音節を表す。図3は各々の音節タイプの母音継続時間の分布であり、図4は子音継続時間の分布である。図3において、母音の継続時間は音節タイプにより異なるが、子音の場合は母音と異なって音節タイプに関係なく類似な継続時間分布を持つ。即ち、母音は音節タイプによって母音の長さが変わるが、子音は音節タイプによって長さが殆ど変化しない。

図3から音節タイプにおける母音の継続時間の関係が
$V > SV, CV > SVC, VC > CSV, CVC > CSVC$ であり、
子音の継続時間の関係は音節タイプに関係なく同じであることがわかる。

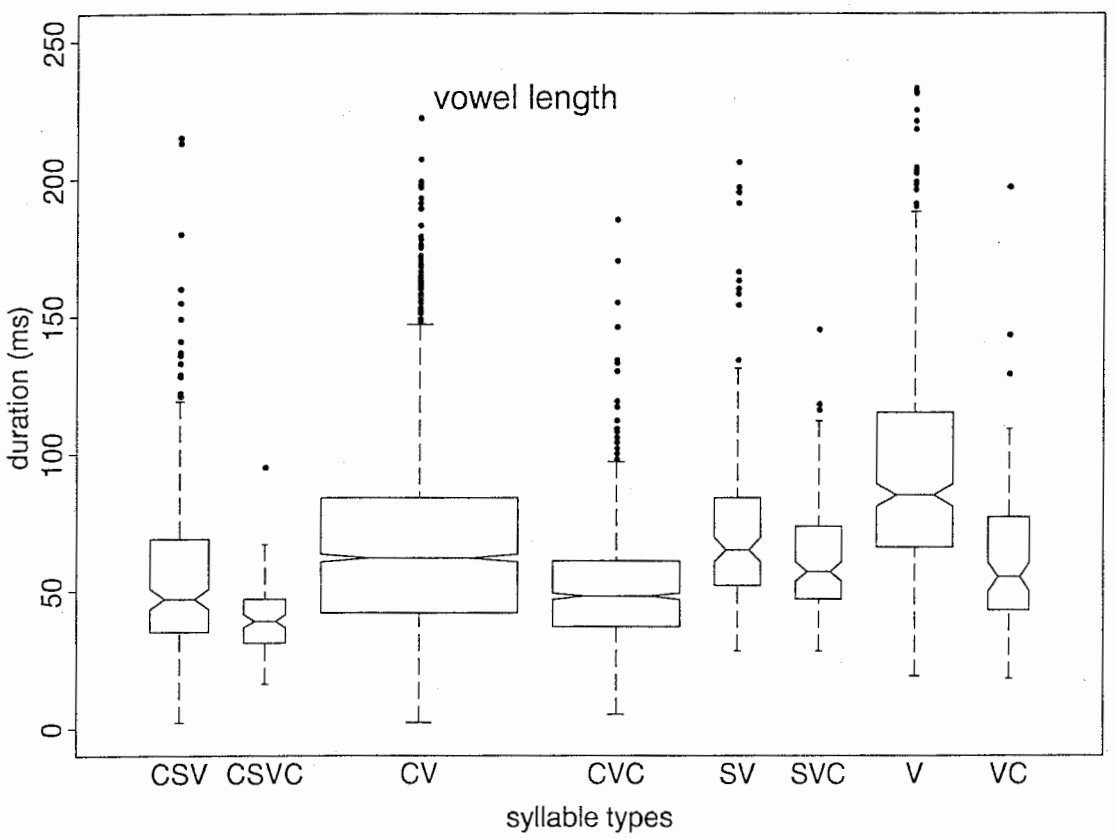

図 3: Variation of vowel duration by syllable type

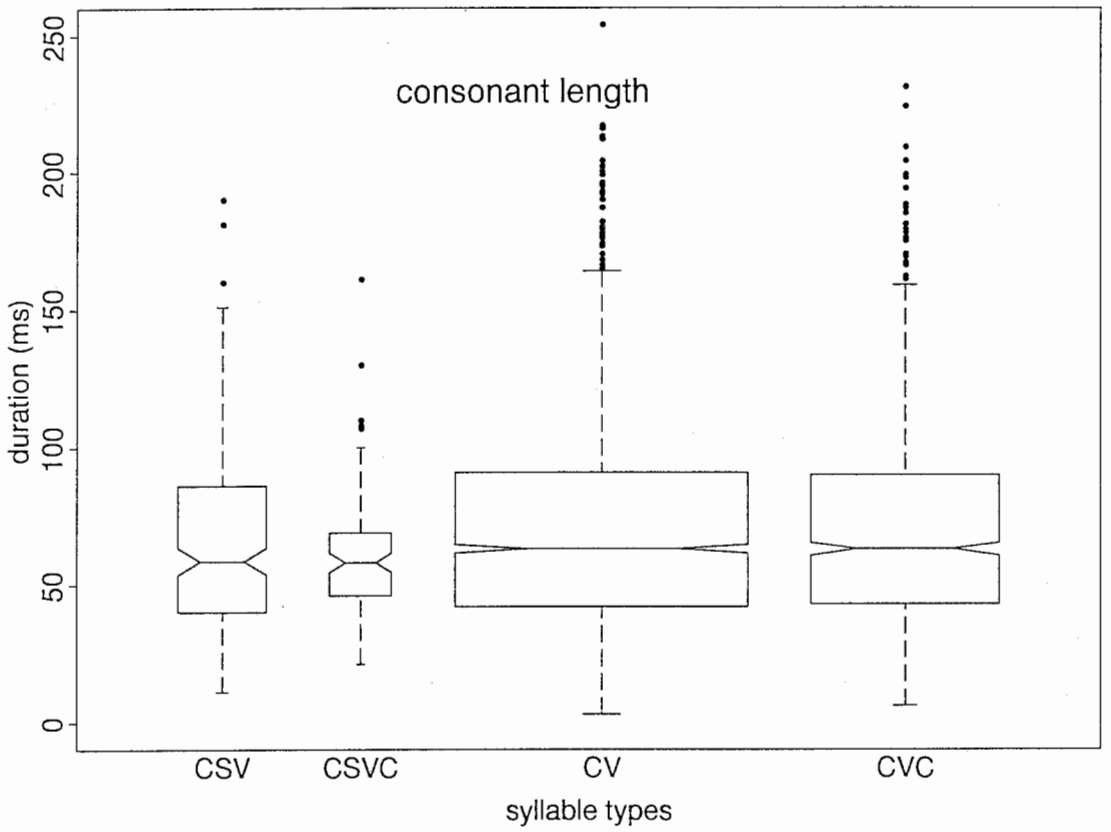

図 4: Variation of consonant duration by syllable type

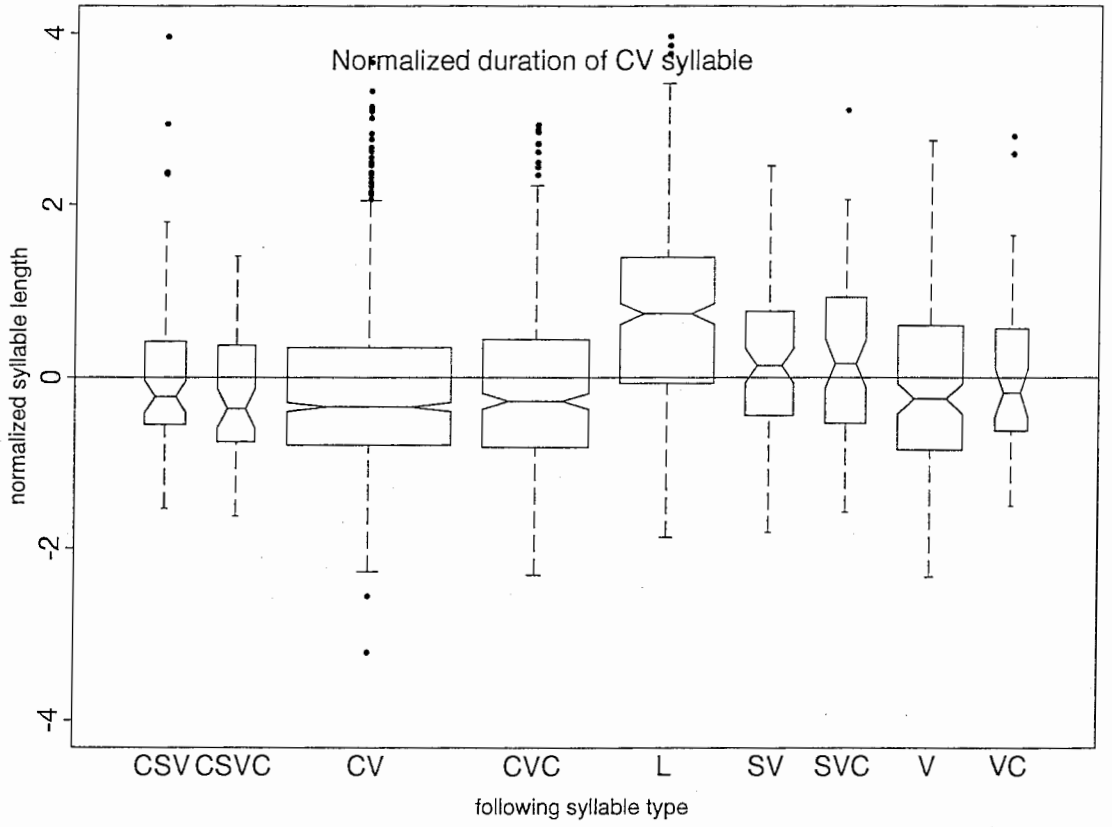

図 5: Variation of syllable duration by following syllable in CV syllable

(3.3.2)　音節継続時間における前後音節の効果

音節間の時間的補償関係は前後の音節タイプによると考えられる。ここでは音声データに多く含まれている CV、CVC 音節に対して、前後音節によって音節の長さがどのように変わるかを調べるために、各々の音節タイプごとの正規化を行ない前後音節タイプで分類する。その結果を図 5、図 6、図 7、図 8に示す。その結果から、ポーズの前の CV 音節の継続時間は他の CV 音節より長くなる傾向があるがポーズの後ろの CV 音節の継続時間は長くならない。このことから CV 音節の場合、文節の最後の音節が最初の音節より長くなる傾向があることが分かる。SV と SVC 音節を除いた全ての音節タイプの前に来る CV 音節の継続時間は短くなるが減少幅がある程度一定である。しかし、CVC 音節の後に来る CV 音節の継続時間は CV 音節の後に来る CV 音節の継続時間より短くなる傾向がある。後に来る音節タイプによる CV 音節の継続時間変化は図 5のようであり、前に来る音節タイプによる CV 音節の継続時間変化は図 6のようである。表 3のようにまとめられる。

また、CVC 音節において、ポーズの前と後ろの CVC 音節の長さは両方ともに長くなるが、ポーズの前より後ろの CVC 音節の長さが長くなる傾向がある。このことから、CVC 音節の場合、文節の最初の音節が一番長くなることが分かる。そして、CV 音節の前に来る CVC 音節の長さが CVC 音節の前に来る CVC 音節の長さより短くなり、CV 音節の後に来る CVC 音節の長さと CVC 音節の後に来る CVC 音節の長さの差は殆どないことが分かる。このことは、

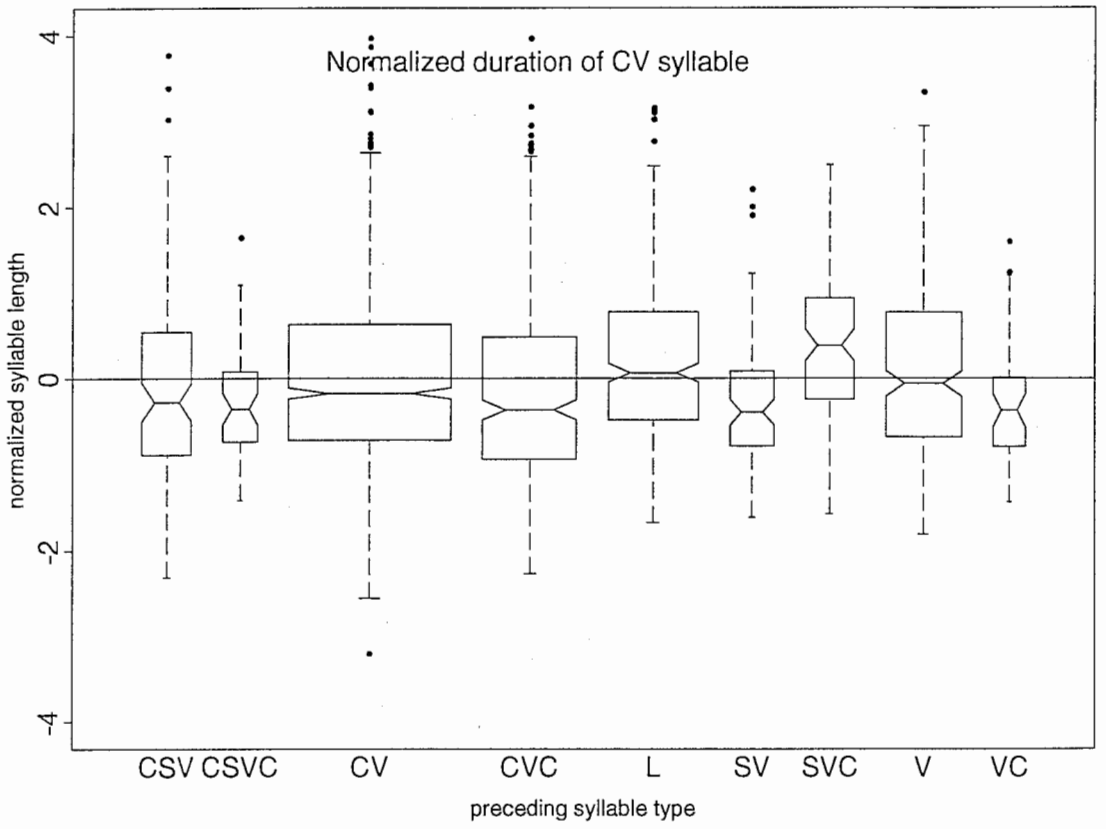

図 6: Variation of syllable duration by preceding syllable in CV syllable

表 3: Summary for the effect of neighboring syllable in CV type syllable

位置	後続音節タイプ	前の音節タイプ
音節継続時間の減少	CV, CVC, $CSVC$, V	$CV > CVC$, SV, VC
音節継続時間の増加	$L(\#)$	SVC

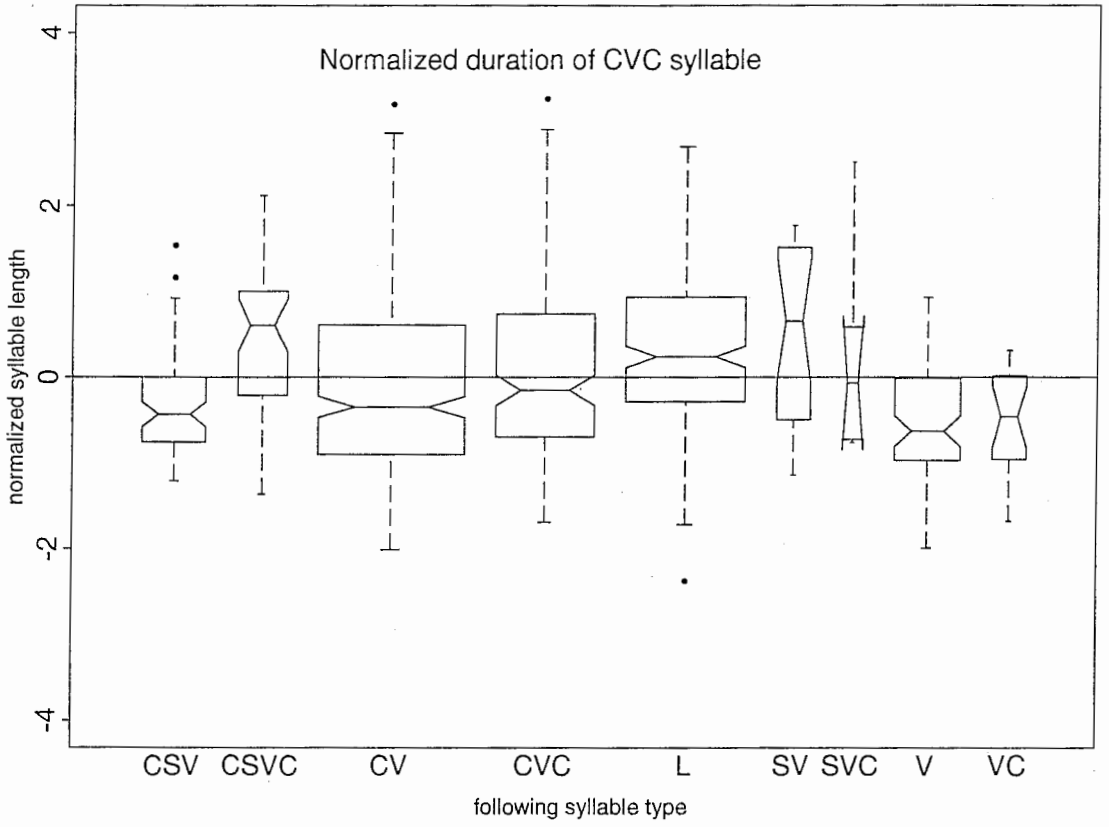

図 7: Variation of syllable duration by following syllable in CVC syllable

CVC 音節において、前の音節の影響よりは後ろの音節の影響が大きいことを表す。CVC の前後音節影響を図 7と図 8に示す。これらの図に現れた／L／はポーズを表す。表 4のようにまとめられる。

(3.3.3)　文節における音節の位置の効果

他言語において文節または句の最後音節の継続時間は最初の音節の継続時間より長くなることが知られている。しかし、韓国語においてはまだ定量的に調べられていない。特に、他言語においては音節のタイプに関係なく句の最後音節の継続時間が長くなると知られている。ここでは、音節タイプに分けて、文節における音節の位置により音節継続時間がどのように変わ

表 4: Summary for the effect of neighboring syllable in CVC type syllable

位置	後続音節タイプ	前の音節タイプ
音節継続時間の減少	$CVC > CV, CSV, V$	CV, CVC, SV
音節継続時間の増加	$L(\#)$	$L(\#)$

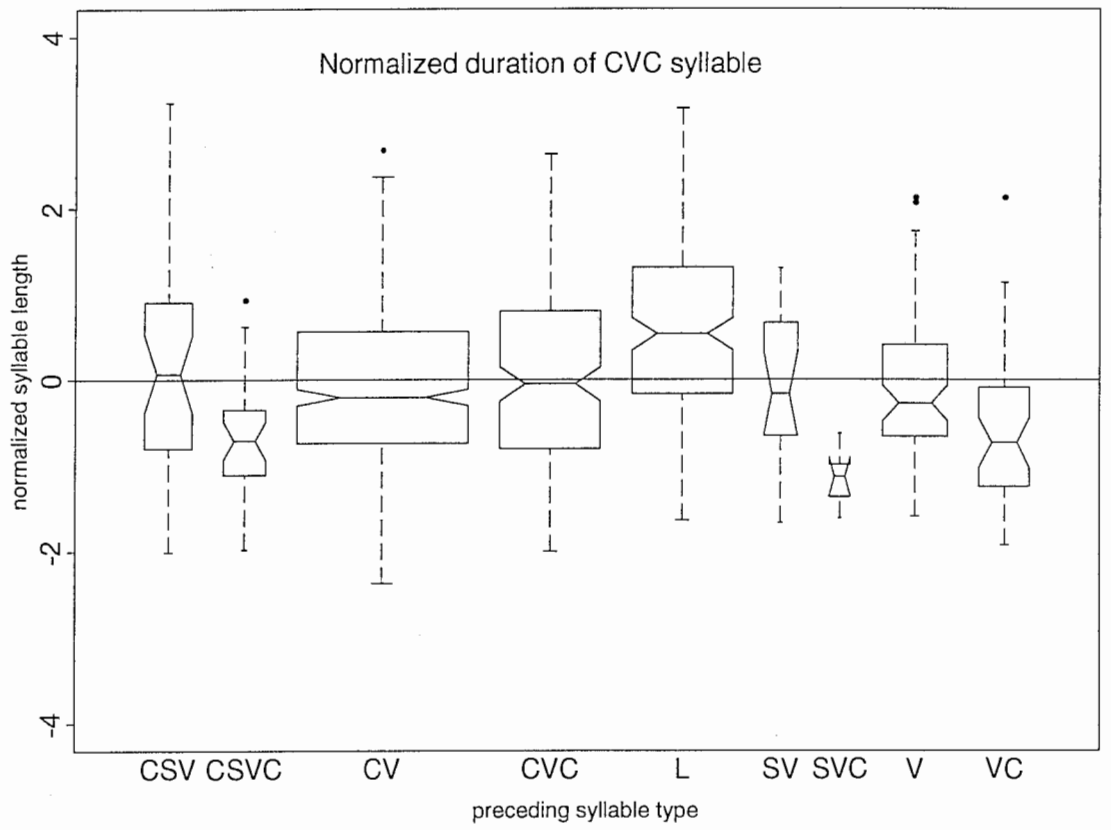

図 8: Variation of syllable duration by preceding syllable in CVC syllable

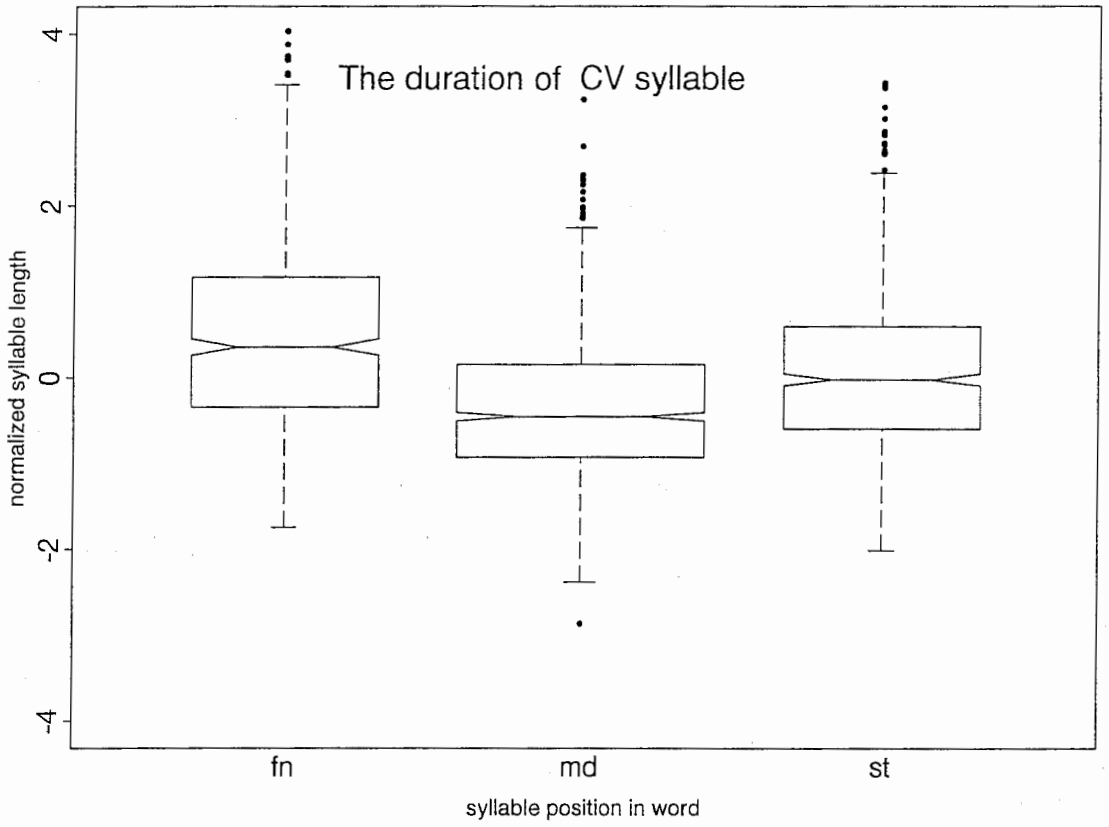

図 9: Normalized duration of CV syllable by syllable position in word

るのかを調べる。そのために、各音節タイプごと Z score による正規化を行ない、文節内の音節の位置別即ち最初、中間、最後に分けて音節の継続時間を調べた結果、CV 音節の場合は図9に示されたように音節の継続時間の長さは
文節の最後の音節 ＞ 最初の音節 ＞ 中間にある音節
のような順である。これは他言語と類似ている。

CVC 音節の場合は図 10のように、他言語とは違って、音節の継続時間長は
文節の最初の音節 ＞ 最後の音節 ＞ 中間にある音節
のような順である。

また、句内の音節の位置による音節の継続時間長も文節内のと同様な継続時間長を持つ。図11に CV 音節、図 12に CVC 音節の句内の位置による正規化された継続時間が示されている。

韓国語の場合は他言語と違って、音節位置により、音節のタイプごとに継続時間長が異なることが分かった。即ち、文節、句共に、CV 音節においては最後の音節が一番長いが、CVC 音節においては最初の音節の継続時間が一番長い。なぜ CVC 音節においては他言語と違って最初の音節の継続時間が一番長いかに対して、韓国語のアクセントは音節構造によって決定されると言う説に結び付けさらに調べられるべきの部分である。

発話テンポに対応する音節の継続時間をモデル化するために、これらの位置に対する音節の継続時間の関係がテンポによりどのように変わるのかを調べる。各音節タイプに対して、文節

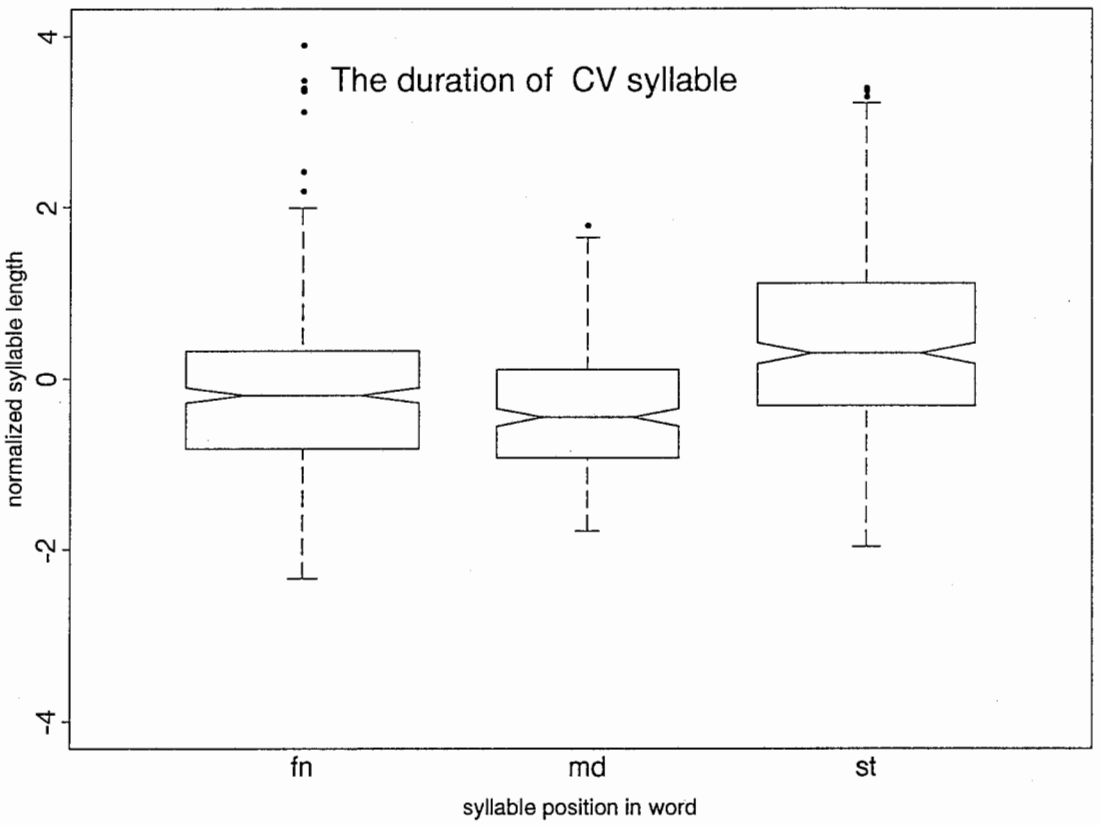

図 10: Normalized duration of CVC syllable by syllable position in word

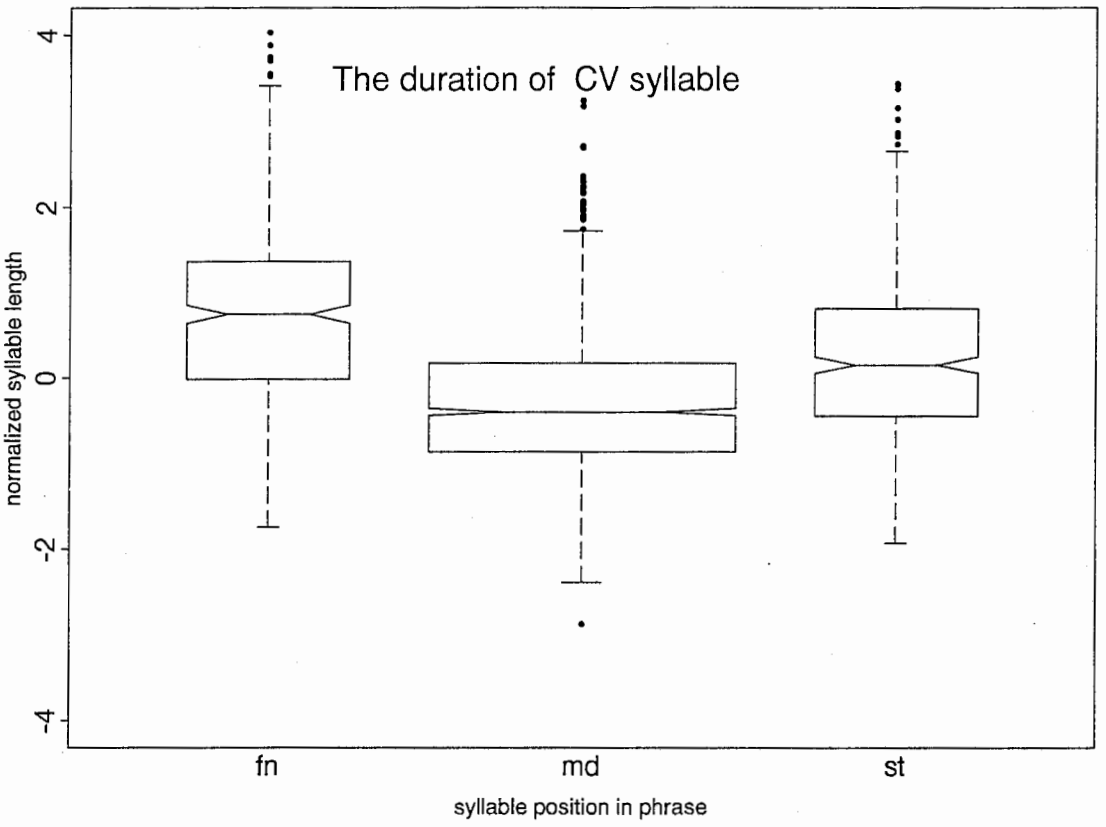

図 11: Normalized duration of CV syllable by syllable position in phrase

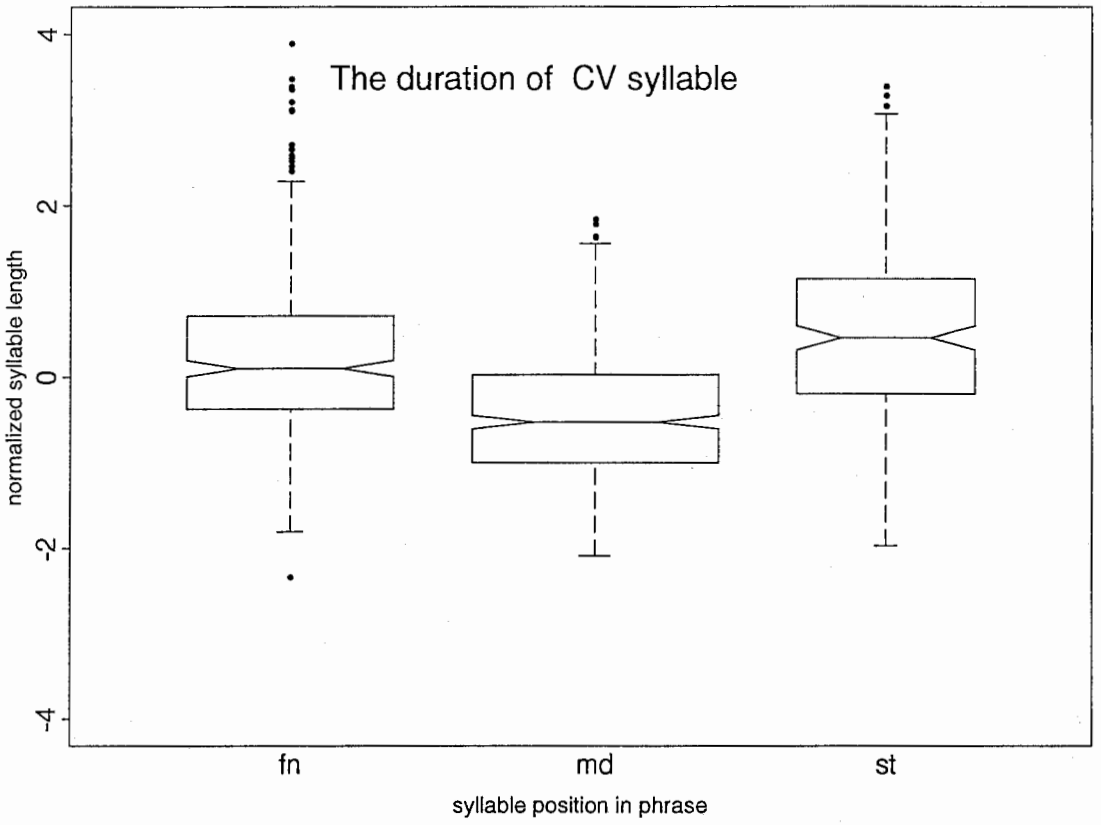

図 12: Normalized duration of CVC syllable by syllable position in phrase

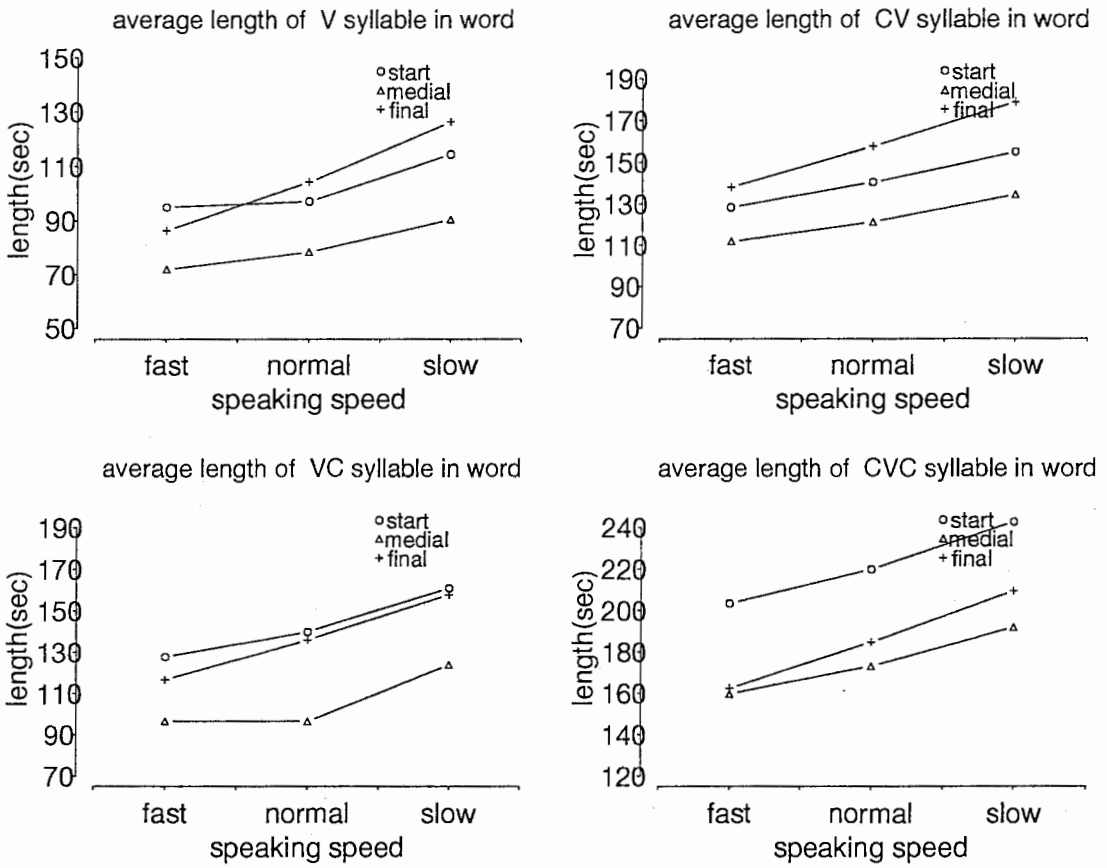

図 13: Variation of syllable duration by syllable position and speaking tempos

内の位置による音節の平均継続時間は図 13のようである。

CV, V 音節において、全てのテンポでの音節継続時間の長さは、V 音節の高速発声を除いて

最終音節 ＞ 最初音節 ＞ 中間の音節

の順であるが、VC と CVC 音節においては

最初音節 ＞ 最終音節 ＞ 中間の音節

のようである。全ての音節タイプにおいて、中間の音節と最初音節との長さの差はテンポに関係なくある程度一定であるが、最後の音節の継続時間はテンポによる変化が大きい。そして, 中間の音節と最初音節において、中速のテンポと高速での変化量が低速での変化量より小さいが、最後の音節の継続時間は中速のテンポと高速での変化量と低速での変化量と同じである。このことは、最後の音節の継続時間が音節タイプに関係なくテンポ変化に一番影響を受けやすいことを表す。最後の音節の弾性が一番大きいとも言える。

(3.3.4)　文節内の音節の数の効果

一般的に、日本語の場合、音韻の継続時間は文節または句内のモーラの数の増加に逆比例すると知られているし、韓国語の場合は音節の継続時間は文節または句内の音節の数に依存すると報告されている。しかし、発声テンポによる音韻または音節の継続時間の変化が調べられ

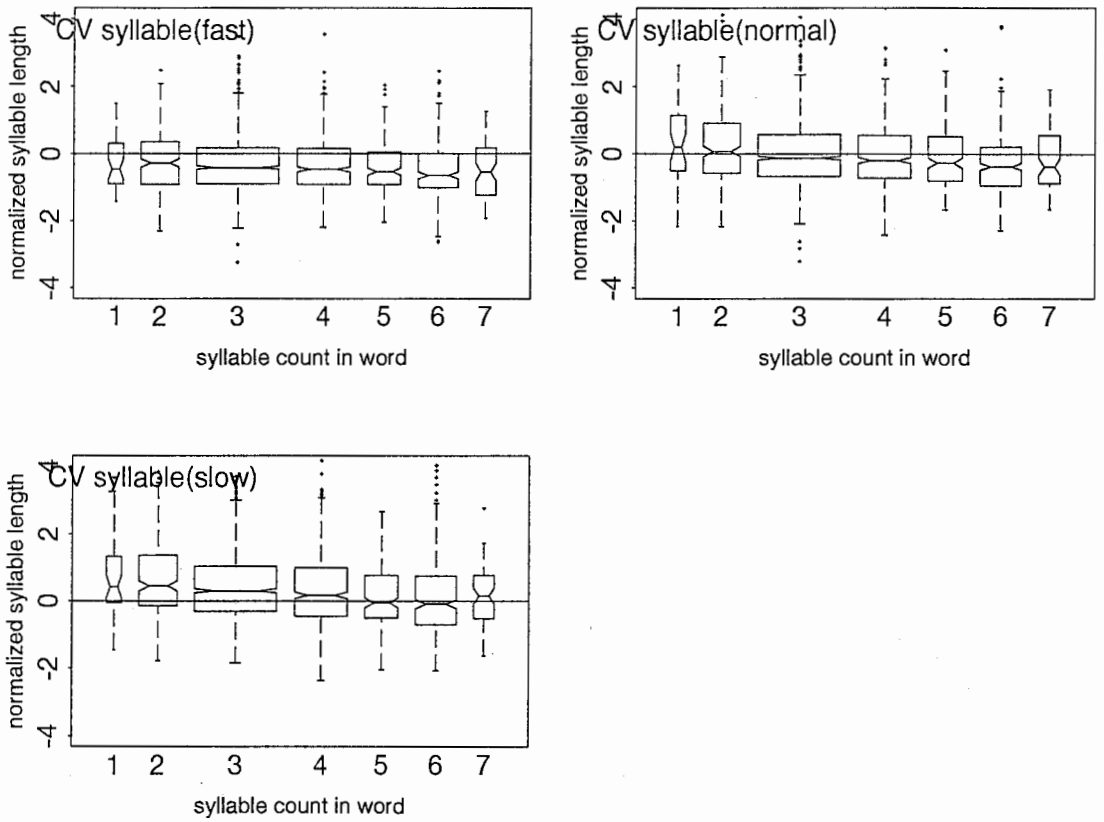

図 14: The effect of syllable count and speaking tempos in consonant duration of CV syllable

ていない。また、韓国語においては、音節の要素、即ち、子音と母音の変化に対して調べられていない。従って、ここでは、発話テンポ別音韻の正規化継続時間を CV、CVC の音節タイプ各々に対して求めたとき、文節内の音節の数による CV 音節の子音の変化を図 14に示す。各々の発話テンポにおいて、子音の継続時間は文節内の音節数の増加により徐々に減少するが、減少幅は小さい。高速テンポの場合は音韻の固有継続時間があるので、音節数の増加による減少幅が他のテンポよりもさらに小さい。文節内の音節の数による CV 音節の母音の変化を図 15に示す。母音の変化は子音変化と似ているが、変化の幅が子音の幅より多少大きい。

また、CVC 音節に対する子音の継続時間の変化は図 16のようであり、母音の変化は図 17のようである。発話テンポに対する変化は CV 音節と似ているが、CVC 音節のは変化幅は CV 音節より大きい。

CV、CVC 音節タイプ共に音節構成要素の継続時間長は音節の数の増加により減少するが、弾性が大きい母音の変化幅が弾性が小さい子音の変化より大きいし、CVC 音節の方が CV 音節の変化幅より大きいことが分かった。また、音韻の固有継続時間のため、高速テンポの変化は他のテンポより小さい。このことは、調音のため最低限の音韻の固有継続時間が必要であるので、縮む幅が伸び幅より小さいことを表す。

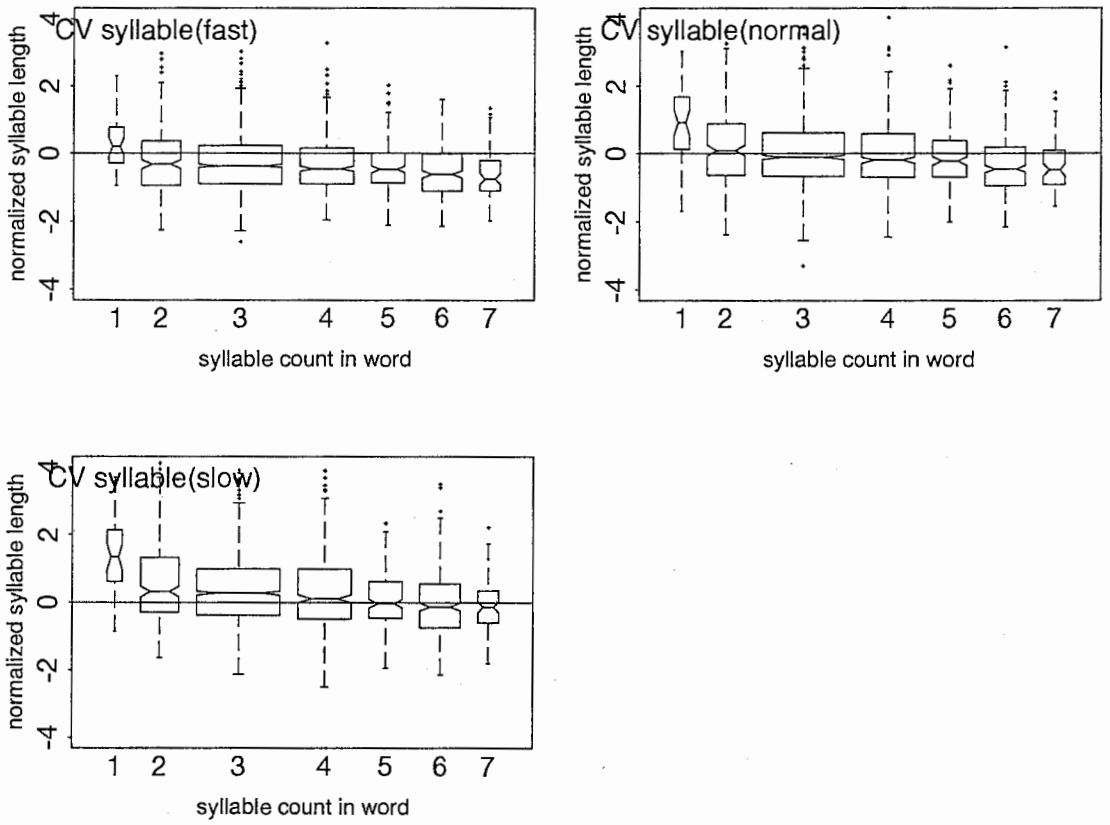

図 15: The effect of syllable count and speaking tempos in vowel duration in CV syllable

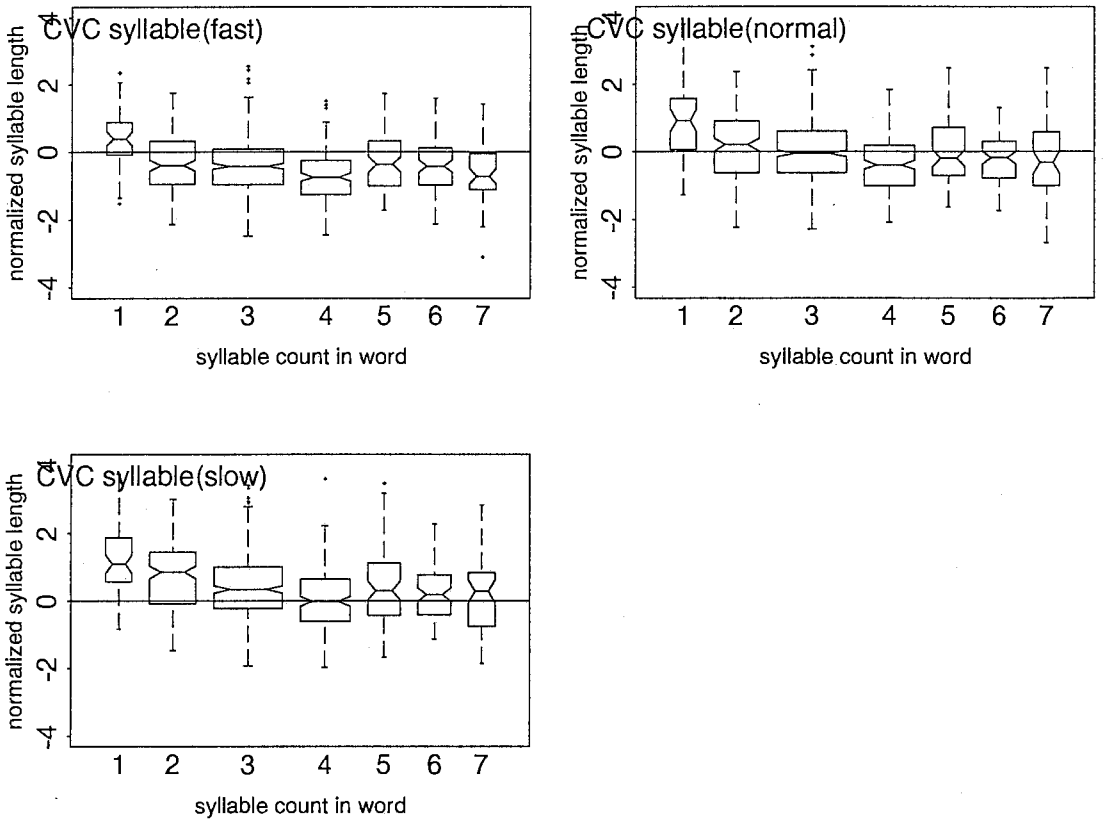

図 16: The effect of syllable count and speaking tempos in consonant of CVC syllable

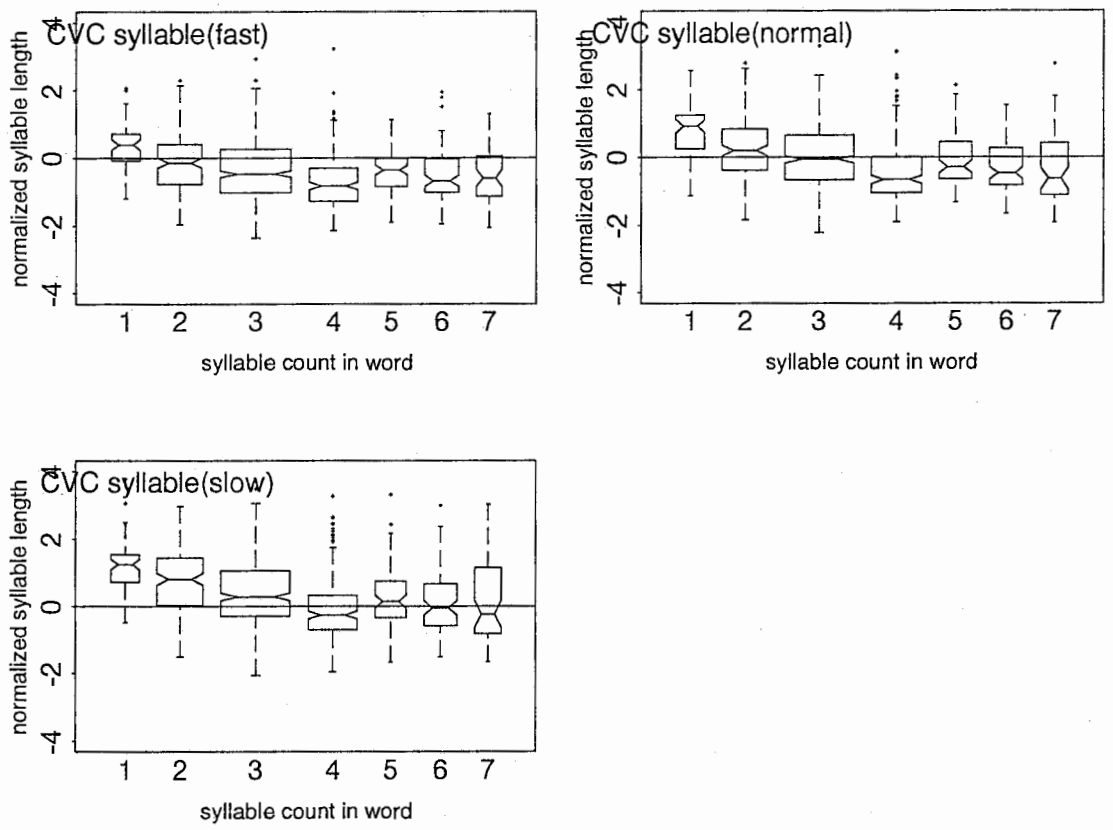

図 17: The effect of syllable count and speaking tempos in vowel of CVC syllable

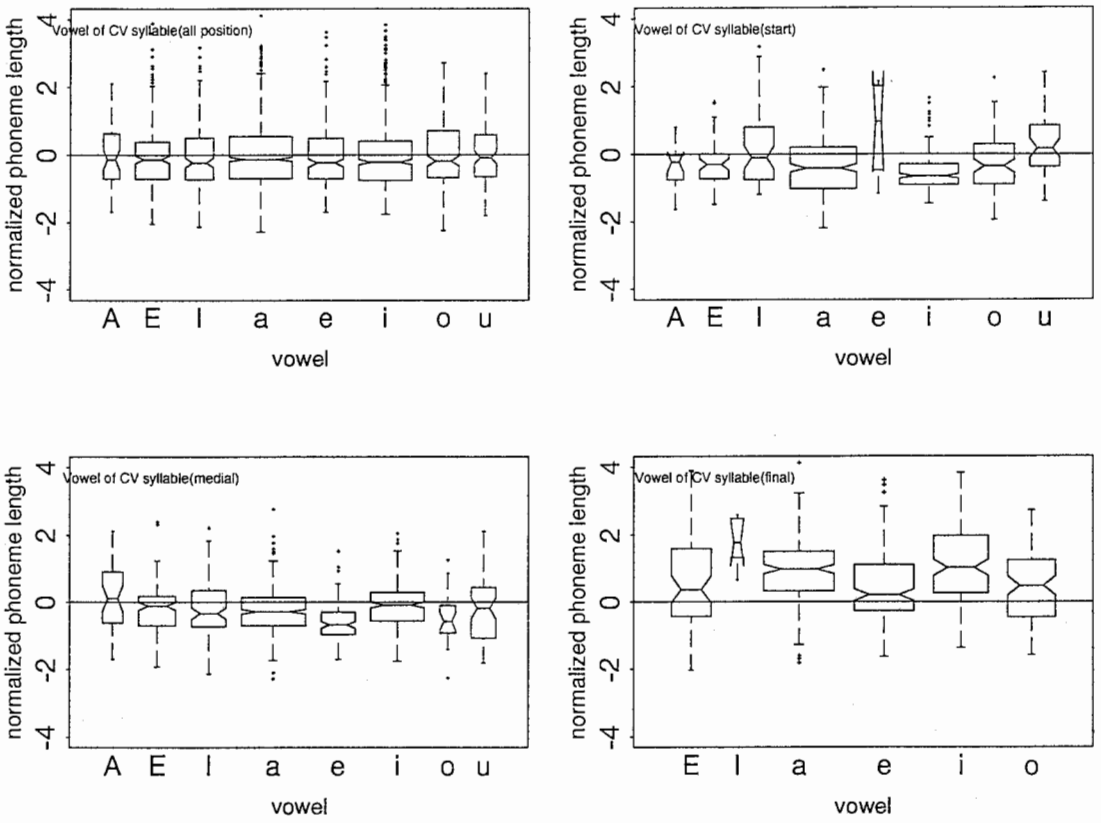

図 18: The effect of syllable position in vowel of CV syllable

3.4 音韻継続時間の特徴

(3.4.1) 音韻継続時間における文節内の音節位置の影響

文節内の音節位置により、音節継続時間は CV、V 音節では最終音節の継続時間が一番長くて中間の音節が一番短かった。そして、VC、CVC 音節では最初音節の継続時間が一番長かった。タイミング制御の要素として採り入れるために、これらの音節を構成する各々の音韻の変化を調べる。各々の音節タイプに対して、音韻の影響を排除するため、各音韻ごとに正規化を行ない、文節内の位置により各音韻の継続時間を求めた。図 18, 図 19に、CV タイプ音節において、文節の内の位置による子音と母音の継続時間を表す。母音では最初と中間の音節では、全般的に短くなり、最終音節では長くなる。しかし、子音では最初の音節が長くなる。即ち、伸び幅は 摩擦音/s/ ＞ 鼻音 ＞ 有気音、硬音、無声子音/g, d, b, z, h/ であり、中間の音節では、有気音と硬音においては平均の長さであるが、他の子音は短くなる。特に、子音 /g,d,b,z/ は有声化のため他の子音よりもっと短くなる傾向がある。即ち、変化の幅は
有音化子音/g, d, b, z, h/ ＞ 無音摩擦音/s/、鼻音、流音 ＞ 有気音、硬音 のようである。最終音節の子音の長さが中間の音節と異なって平均の長さを保つ。CVC 音節に対しても類似であるが、母音の長さは大きく変わらない。

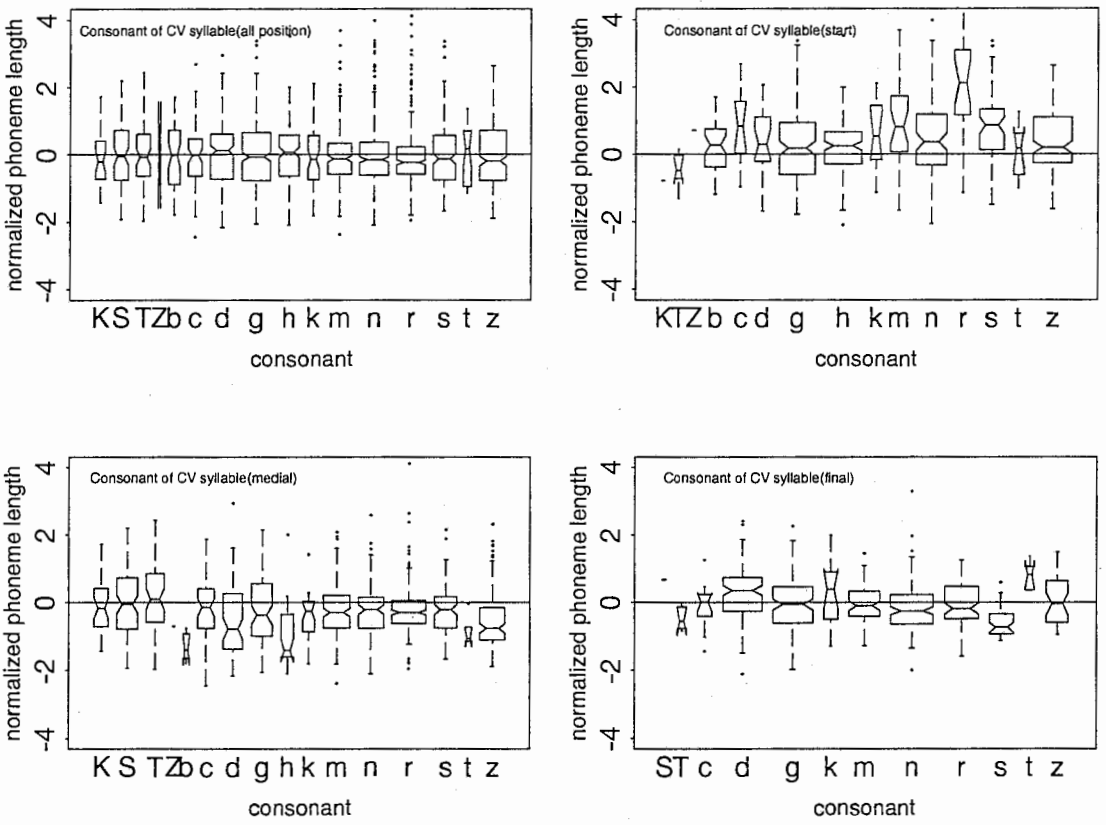

図 19: The effect of syllable position in consonant of CV syllable

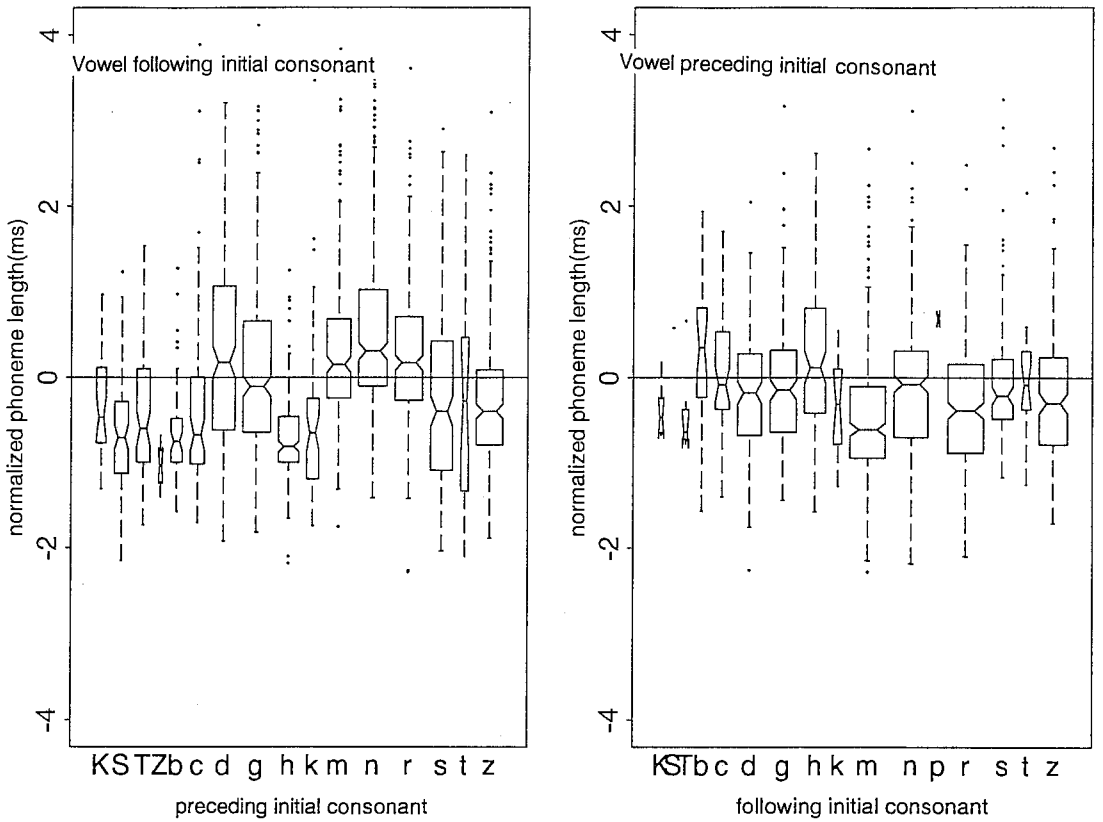

図 20: The effect of neighboring consonant for vowel duration in CV syllable

(3.4.2) 音韻継続時間における隣接音韻の影響

日本語に対して、モーラ タイミングから時間的補償効果が良く知られている。文献 [5] では韓国語に対して、母音の前後音韻の両方の組合せの効果に対して調べタイミング制御に用いた。しかし、正確な音韻の特徴を得るためには、前の音韻、後ろの音韻に分離した各々の効果を調べる必要がある。ここでは CV 音節と CVC 音節における各音韻の前の音韻と後ろの音韻の影響を分離して調べる。図 20左側に、CV 音節において、母音継続時間に対する前の子音の効果が示されている。この図から、母音の継続時間は前の子音により著しく変わる。子音の種類よる母音の長さは次のようであることが分かる。

- 鼻音、流音 > 有声破裂音 > 有気音、硬音、摩擦音、破擦音

また、CV 音節において、母音継続時間に対する後ろの子音の効果が図 20右側に示されている。ここで、全ての後ろの子音により、母音の継続時間が全般的に短くなる傾向がある。特別な子音により、母音の継続時間が大きく変化しないことである。このことから、CV 音節において、母音の継続時間は音節フレームにより、CV 音節の子音と母音の間の調音結合が他音節の子音との調音結合より強いので、前の子音の影響が後ろの子音の影響より大きいと考えられる。図 20左側では前の子音が長くなると母音が短くなって、子音が短くなると母音が長くなる現状が見られる。これは、音節の固有の継続時間があって、音節内での音韻の相互補償のた

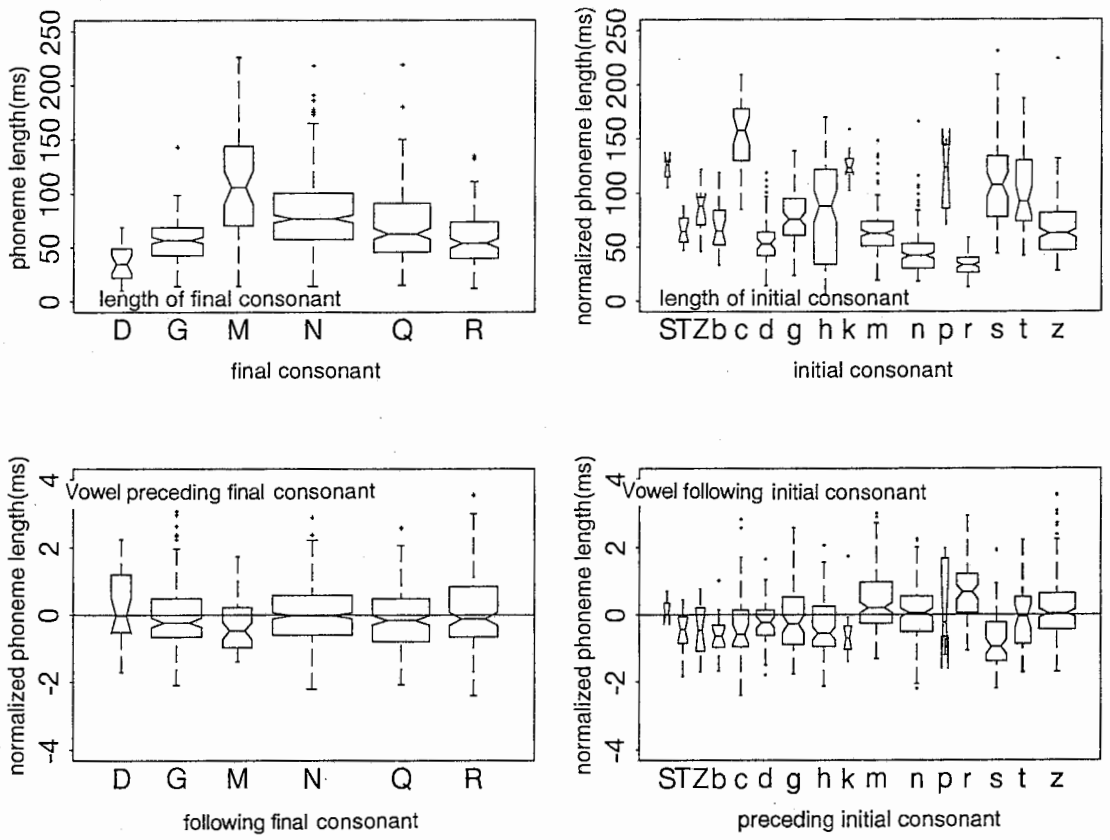

図 21: The effect of neighbouring phoneme for vowel duration in CVC syllable

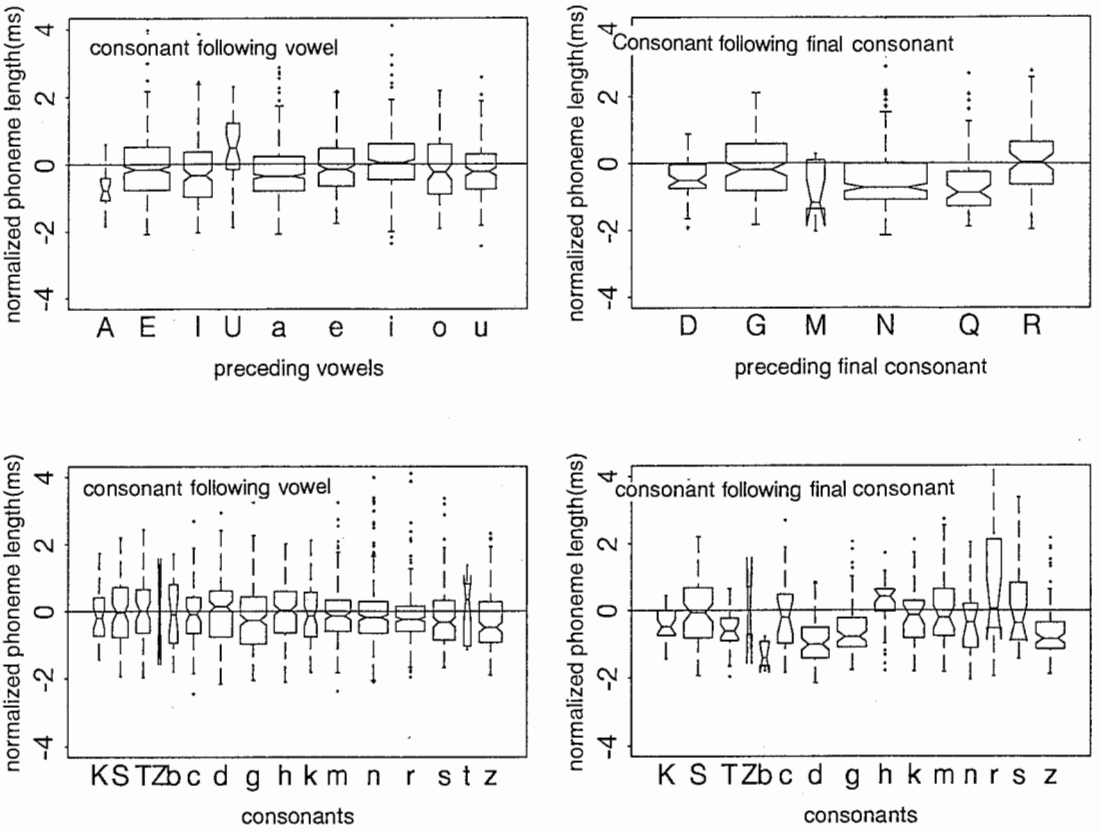

図 22: The effect of neighboring phoneme in vowel of CV syllable

めであると考えられる。

CVC 音節において、母音の前の子音の継続時間長に対するその後ろの母音の正規化継続時間長の変化を図 21右側に、後ろの子音の継続時間長に対するその前の母音の正規化継続時間長の変化を図 21左側に示す。母音継続時間長に対する前の子音の影響は CV 音節と類似であるが、後ろの子音、即ち、終声の影響は後ろの子音の音韻とは関係なく一定である。このことから、母音の継続時間長は後ろの音韻より前の音韻の影響が大きいことが分かる。

また、子音継続時間長に対する前の音韻の効果は図 22に示す。前の音韻が母音の場合、図 22左側に示すように、子音の継続時間長の変化は殆どないが、前の音韻が終声の場合（図 22右側）、鼻音の終声の後ろの子音が短くなる。これは、子音の有声音化が強く現れたことを意味する。

4 回帰木による音韻継続時間のモデル化

4.1 制御要素

観測データからその制御規則を推測するための統計モデルとして、説明変数空間を逐次ツリー分割することで、カテゴリー間の依存関係などの分布の非線形性を表現できる回帰木を用いる。このような統計的方法では特徴セットの選択が非常に重要である。ここで使われた特徴セットは前に分析してきた音韻の継続時間長を変化させる重要な要因である。従って、これらの要因を用いてタイミング制御のモデリングを行なう。音節継続時間長予測のための重要な要因は次のようである。

- 音節コンテキスト：予測する音韻を含む音節（音韻継続時間長予測）、予測する音節、予測する音節の左右 1 つ音節（音節継続時間長予測）
- 音韻コンテキスト：予測する音韻、予測する音韻の左右 2 つ音韻（音韻継続時間長予測）
- 文節内の位置：文節の最初と最後と中間の音節、文節内の音節の数（音韻継続時間長予測、音節継続時間長予測）
- 句内の位置　：句内の最初と最後と中間の音節、句内の音節の数（音韻継続時間長予測、音節継続時間長予測）

ここで、各音節は 9 種類の音節タイプクラス、即ち、V、SV、CV、CSV、VC、SVC、CVC、$CSVC$、SIL に分類される。また、各々の音韻は調音様式と位置により、子音は 13 種類、plosive lenis, plosive aspiration, nasal bilabial, nasal alveolar, voiced nasal bilabial and alveolar coda, voiced liquid coda, voiced nasal velars coda, unvoiced coda, fricative alveolar, fricative glottal, plosive and affricate fortis, fricative and affricate fortis, semivowel
母音は 4 種類、low vowel, mid vowel, high vowel, compound vowel
合計 17 種類に分類される。文節と句内の位置は 3 種類、即ち、始まり、中間、終りの音節に分類する。テキストで表した回帰木の例を付録に示す。

4.2 予測結果

7 人の話者各々に対して、中速テンポの 16 発話（文）の中で 15 発話（文）が回帰木を生成するための訓練データとして用いられ、残り 1 つの文の各々のテンポに対する発話はテストデータとして用いられた。全ての 16 文に対して同様の予測が行なわれた。その結果を図 23に示す。この結果によれば、話者によって異なるが、女性の方が男性より良く予測できた。また、高速発声より低速発声の方が良く予測できた。図 24には音節継続時間長の予測結果である。ここで音節の継続時間長の方が音韻の継続時間長より良く予測できた。平均多重相関は表 5のようである。

音声データ内の子音と母音の継続時間長分布は図 25のようである。子音と母音の予測の残差は図 26、図 27のようである。これらの図から子音、母音共に 75 女性話者 1 に対する文 1 の音韻、音節継続時間長の予測値と観測値間の相関は図 28、図 29のようである。従って、こ

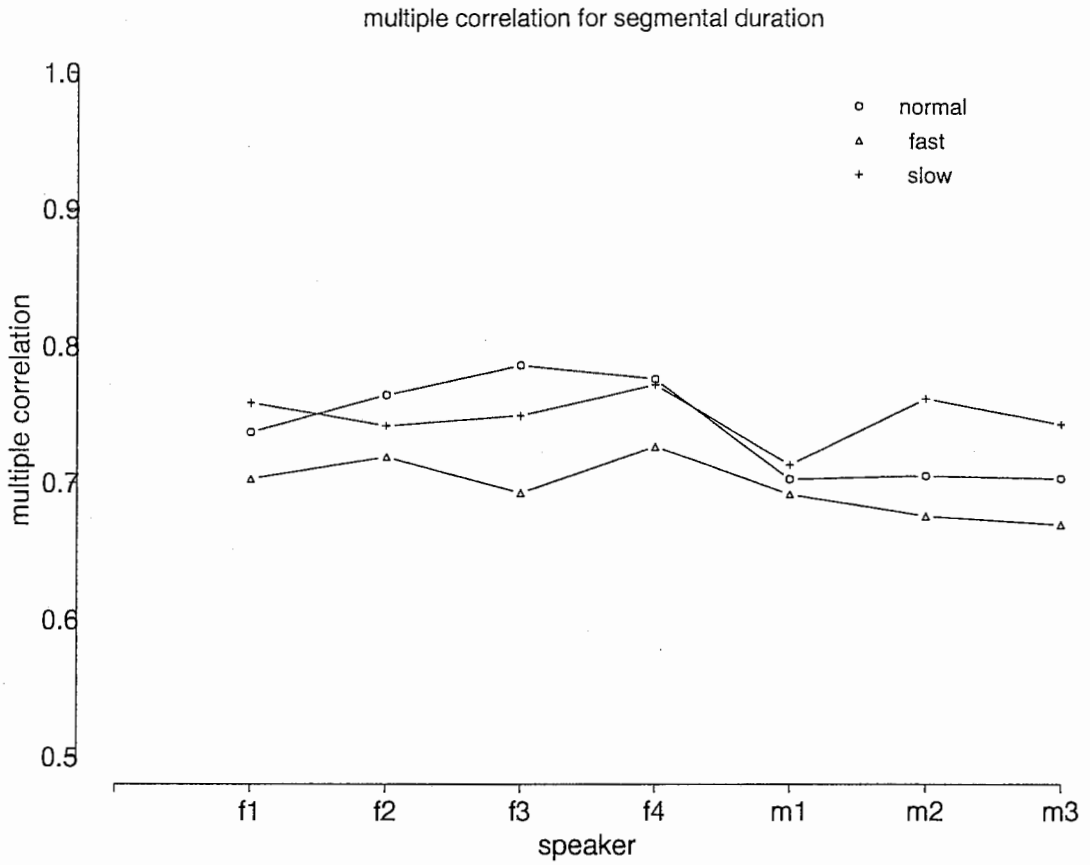

図 23: Multiple correlation between predicted and observed duration(segmental)

表 5: Multiple correlation between predicted and observed duration

tempo	segmental level	syllable level
normal	0.72	0.86
fast	0.69	0.83
slow	0.74	0.88

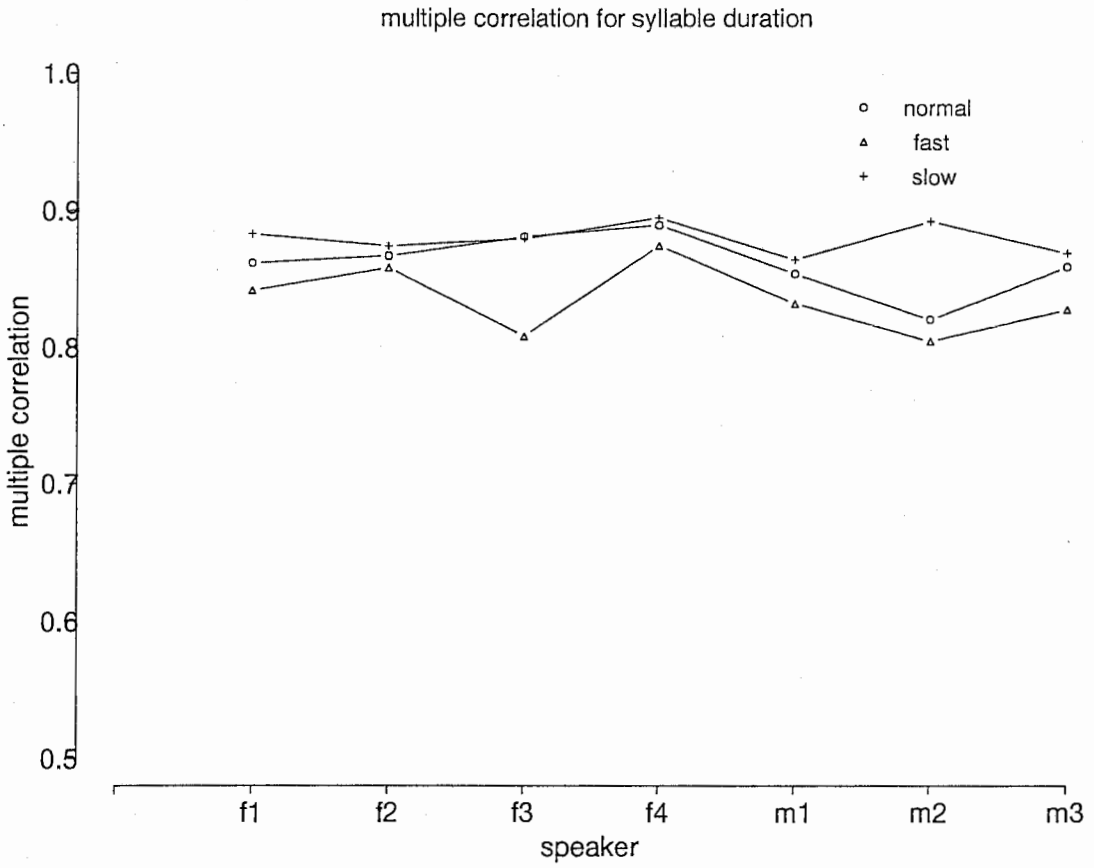

図 24: Multiple correlation between predicted and observed duration(syllable)

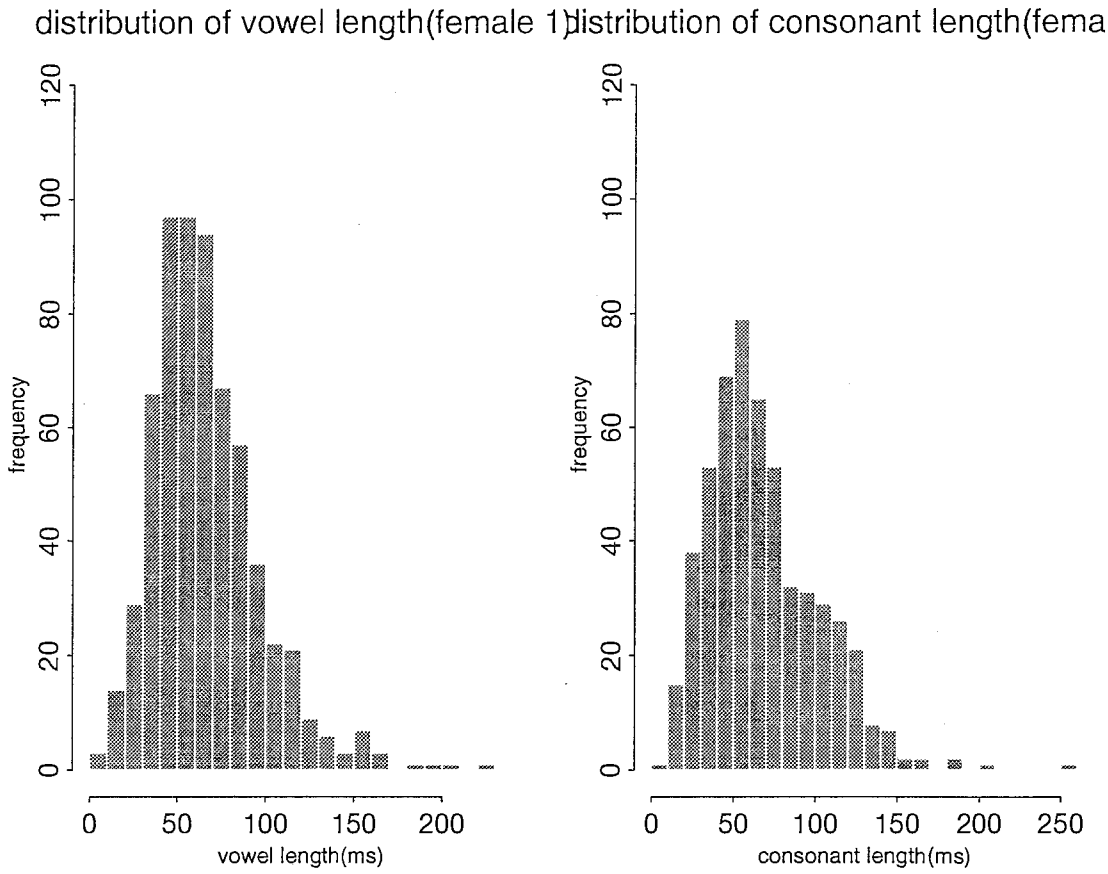

図 25: Distribution of segmental duration(speaker:female 1)

こで生成された音韻継続時間長モデルは妥当であることを表す。図 30に音韻別残差の分布を示す。このモデルに品詞情報など文法的情報を加えることによって、より良い予測が可能である。

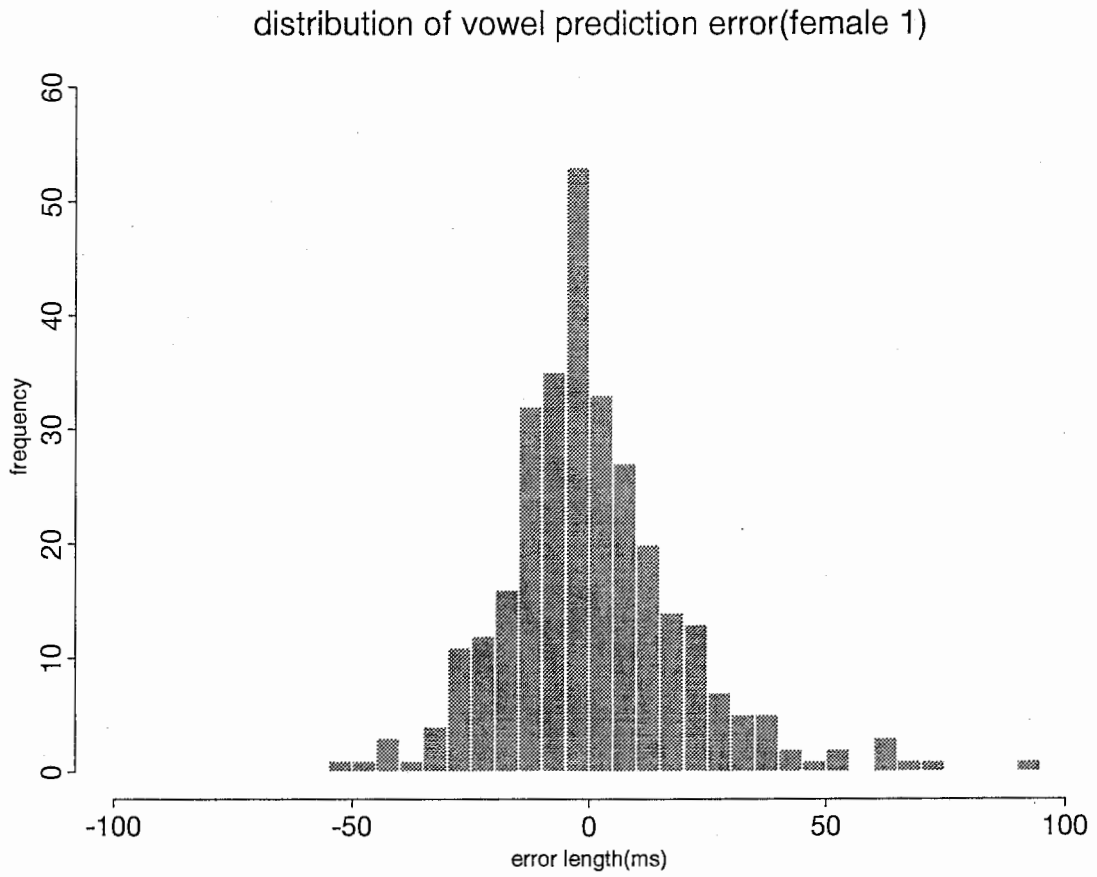

図 26: Distribution of vowel prediction error by regression tree

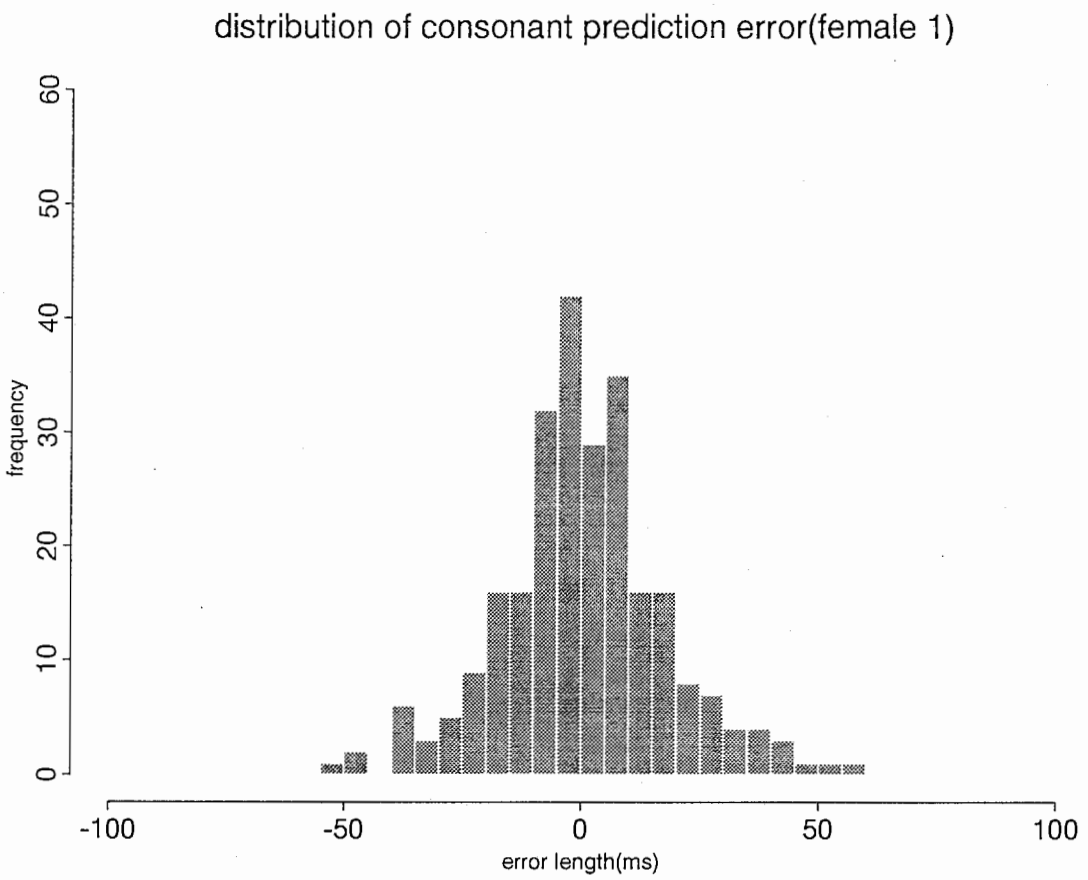

図 27: Distribution of consonant prediction error by regression tree

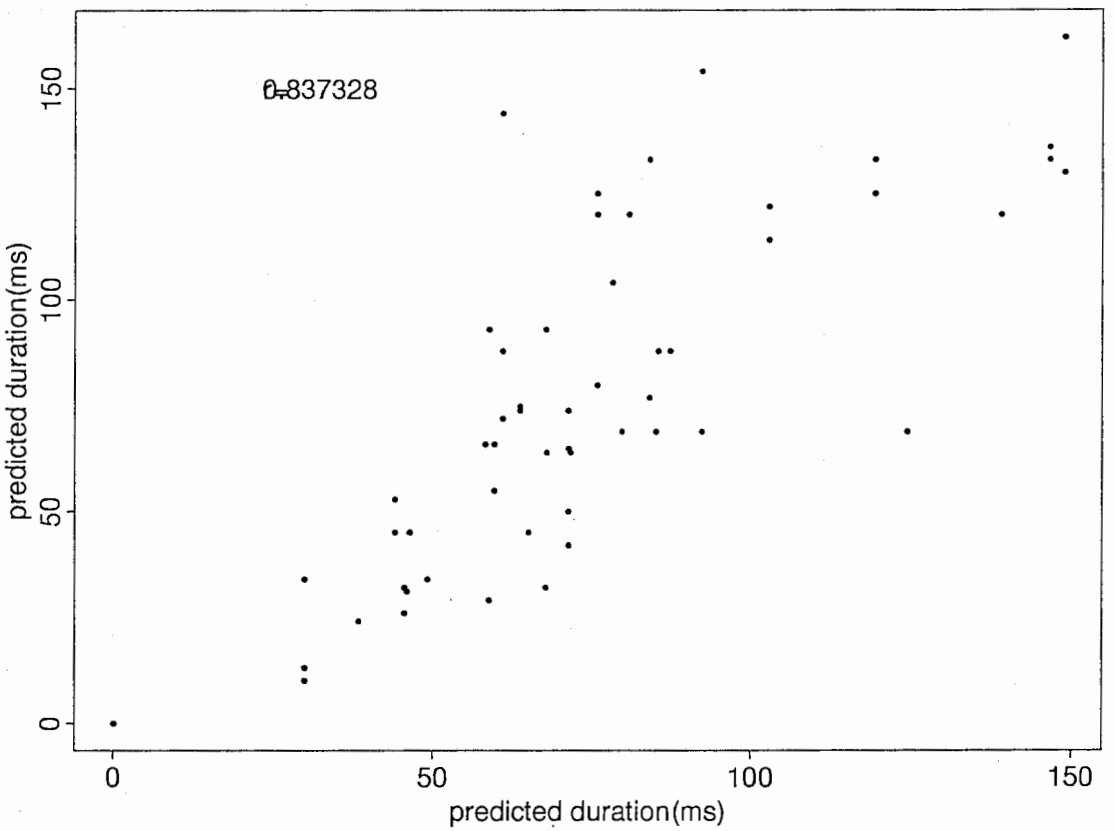

図 28: Correlation between predicted and observed duration(segmental)

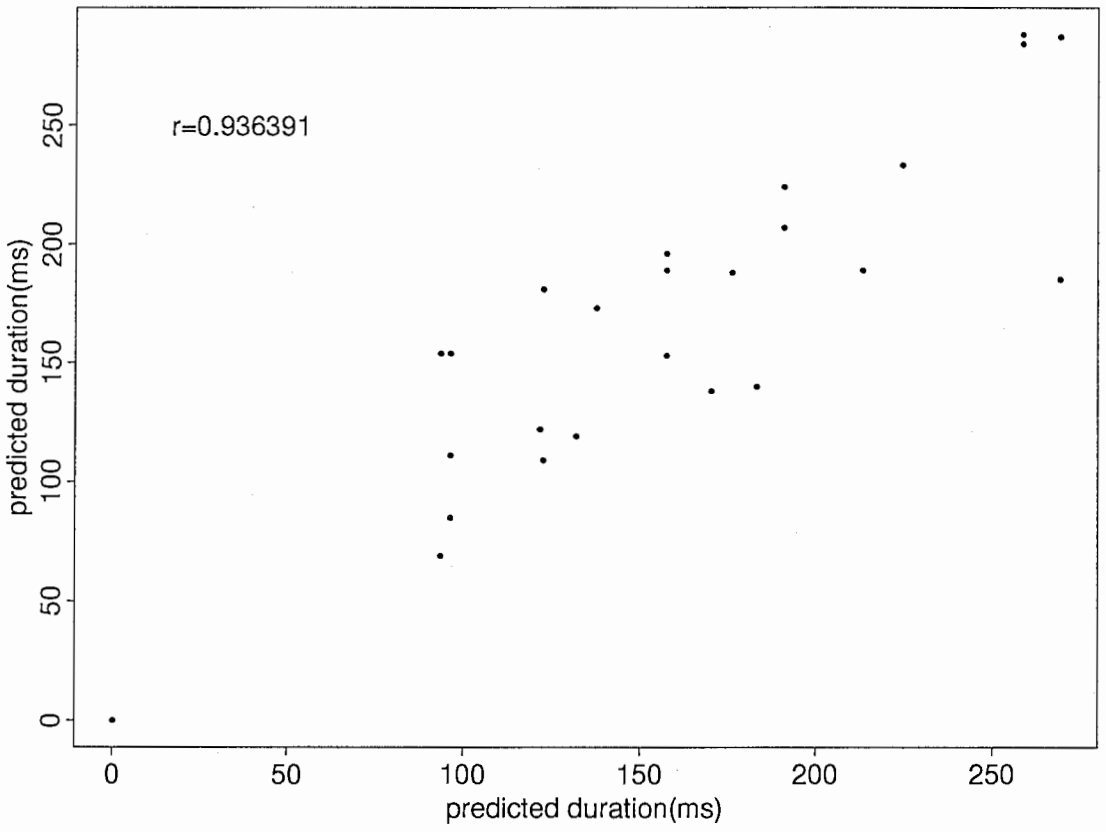

図 29: Correlation between predicted and observed duration(syllable)

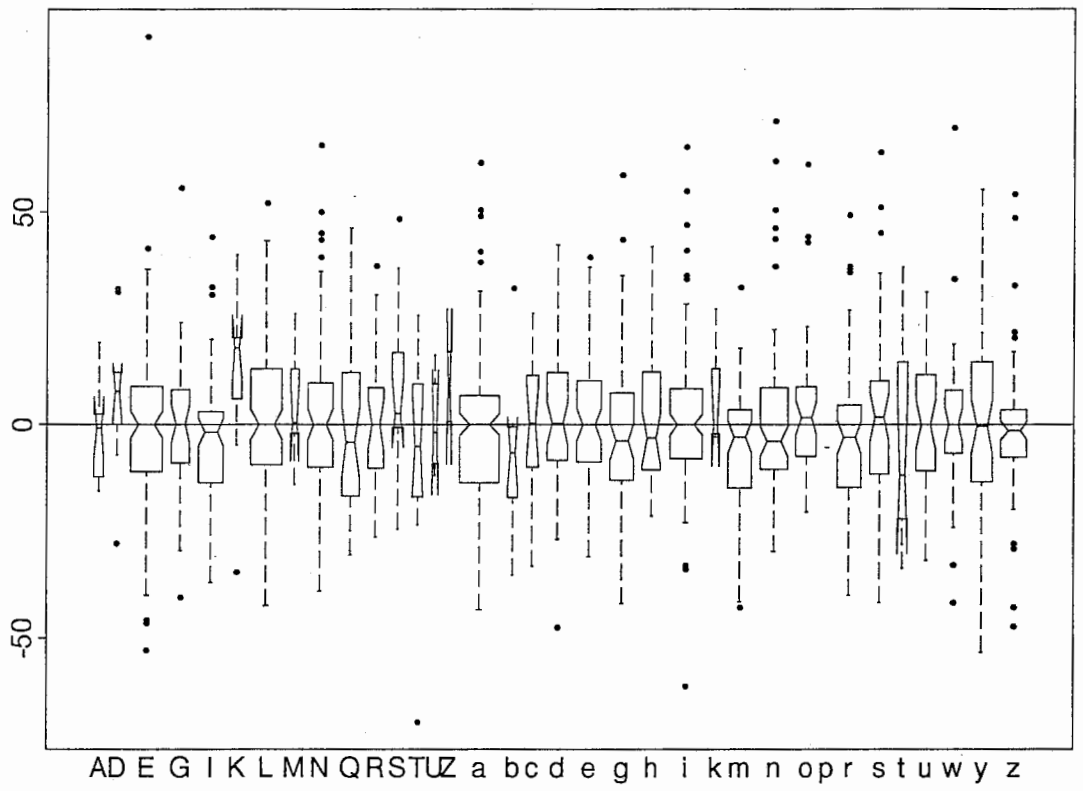

図 30: Distribution of predicted error for each phoneme

5 おわりに

韓国語の音声合成において自然な音声を合成するために、韓国語のタイミング特徴を統計的に分析した。その結果、1）発話テンポはポーズと文節または句の境界音節の継続時間に依存する。 2）音節タイプごとに固有の母音継続時間を持つ。 3）CVC 音節では文節または句の最初音節で長くなり、CV 音節では最終音節で長くなる。 4）音節、音韻の継続時間は文節または句内の音節数に逆非例する。 5）母音の継続時間は次に来る子音より前の子音の影響が大きいことが明らかになった。また、これらの特徴セットを回帰木に用いて、音韻と音節の継続時間をモデリングした。この研究で生成したモデルを使って継続時間の予測を行ない、その有効性を確認した。今後は現在モデルの制御要素に品詞情報及び文法的情報を加え、予測の改善を図る。

謝辞

ATRで研究する機会与えて下さいました山崎社長に心から感謝致します。並びに、色々支援して下さいました樋口室長、ローケンキムさん、また、貴重な討論頂いたニックさんに心から深謝致します。ご親切な第2研の皆様にこの場を借りて感謝の気持ちを伝えます。

参考文献

[1] D.H.Klatt, "Linguistic uses of segmental duration in English: Acoustic and perceptual evidence," *Journal of the Acoustical Society of America*, 59, 5, pp1208-122 1,1976.

[2] H.B.Lee *et al.*, "An experimental phonetic study of speech rhythm in standard Korean," *Proc. of ICSLP*, pp.1091-1094,1994.

[3] N. Kaiki, K. Takeda and Y.Sagisaka, "Linguistic propertyes in the control of segmental duration for synthesis," *Talking machines : Theories, Models, Designs*, pp255-263,1992.

[4] Jan P.H. van Santen, "Deriving text-to-speech drations from natural speech", *Talking machines : Theories, Models, Designs*, pp275-285,1992.

[5] S.H.Kim and J.C.Lee, "Korean text-to-speech system using time domain-pitch synchronous overlap and add method", *International conf. on SST*, vol.2, pp315-320, 1994.

[6] M.D. Riley, "Tree-based modelling of segmental durations," *Talking machines : Theories, Models, Designs*, pp265-273, 1992.

[7] W.N. Campbell, "Syllable-based segmental duration," *Talking machines : Theories, Models, Designs*, pp211-223, 1992.

[8] 岩橋 直人、匂坂芳典、"空間分割型数量化法による音声制御の統計モデリング," 日本音響学会秋季研究発表会講演論文集, pp237-238, 1992.

[9] 岩田和彦、三留幸夫、"発話テンポに依存しない韻律構造のモデル化," 電子情報通信学会春季大会講演論文集, pp1-486 - 1-487, 1994.

[10] 匂坂芳典、"日本語音声の韻律的特徴とその計算モデル," 日本音響学会秋季研究発表会講演論文集, pp295-298, 1994.

6　回帰木の例

```
node), split, n, deviance, yval
      * denotes terminal node

   1) root 1477 2006000.0  60.25
     2) phon:,fcu,hv,li,nb,sv 570  404300.0  36.03
       4) phon: 107       0.0   0.00 *
       5) phon:fcu,hv,li,nb,sv 463  233300.0  44.35
        10) nxtphon:a2,af,fc1,fc2,fcu,hv,li,lv,mv,na,nb,sv 380  123300.0  40.13
          20) prephon:a2,fa,fc3,fg,fo,ln,sv 97   20190.0  30.90
            40) nxtphon:af,fc1,fcu,hv,li,lv,mv,na,sv 78   12770.0  27.68 *
            41) nxtphon:a2,fc2,nb 19    3304.0  44.11 *
          21) prephon:,a1,af,fc2,hv,li,lv,mv,na,nb 283   91990.0  43.30
            42) pre2ph:a1,fa,fc1,fc2,fc3,fg,fo,ln,na,nb,sv 128   23270.0  37.58
              84) phon:li,sv 59    7533.0  31.24 *
              85) phon:fcu,hv,nb 69   11340.0  43.00 *
            43) pre2ph:,a2,af,fcu,hv,li,lv,mv,uv 155   61070.0  48.03
              86) nosp<4.5 71   36290.0  53.66
               172) nxtphon:a2,fc1,fc2,hv,li,lv,mv,nb,sv 63   19610.0  49.84 *
               173) nxtphon:af,fcu,na 8    8520.0  83.75 *
              87) nosp>4.5 84   20620.0  43.26
               174) nxt2ph:af,fc2,fcu,fg,fo,hv,ln,lv,mv,nb 59   12680.0  38.86 *
               175) nxt2ph:,a2,fc3,li,na,sv 25    4098.0  53.64 *
        11) nxtphon:,a1,fa,fc3,fg,fo,ln,uv 83   72300.0  63.66
          22) prephon:fa,fo,hv,li,ln,lv,mv,nb 61   20960.0  53.15
            44) pre2ph:a2,fc2,fcu,uv 7    3147.0  28.71 *
            45) pre2ph:,af,fa,fc3,fg,hv,li,ln,lv,mv,sv 54   13100.0  56.31 *
          23) prephon:,a2,af,fc2,na,sv 22   25890.0  92.82
            46) nxtphon:fa,fc3,ln 14   13210.0  79.86 *
            47) nxtphon:,fg,fo 8    6214.0 115.50 *
     3) phon:a1,a2,af,fa,fc1,fc2,fc3,fg,fo,ln,lv,mv,na,uv 907 1057000.0  75.47
       6) phon:af,fc1,fc2,fc3,ln,lv,mv,na 760  603900.0  68.10
        12) prephon:a1,a2,af,fa,fc1,fc2,fc3,fg,fo,sv 277  155500.0  55.37
          24) nxtphon:a1,fc2,fc3,fcu,fg,fo,hv,li,ln,lv,mv,na,nb,uv 221   96700.0  51
            48) nxt2ph:,af,fa,fc2,fcu,fg,fo,li,ln,mv,na,nb,sv 149   53260.0  47.30
              96) prephon:fa,fc1,fg,fo 26    2612.0  35.00 *
              97) prephon:a1,a2,af,fc2,fc3,sv 123   45890.0  49.89
               194) stype:CSV,CSVC,CV,VC 73   20960.0  45.22
```

```
          388) phon:ln,lv,mv,na 63   12580.0  42.05 *
          389) phon:af 10    3754.0  65.20 *
        195) stype:CVC,SV,SVC 50   21000.0  56.72
          390) nosp<7.5 44   13860.0  54.07 *
          391) nosp>7.5 6    4555.0  76.17 *
      49) nxt2ph:a1,a2,fc3,hv,lv 72   37150.0  58.68
        98) pre2ph:li,ln 7    1841.0  33.86 *
        99) pre2ph:,fc2,fc3,fcu,hv,lv,mv,nb 65   30530.0  61.35
         198) nxtphon:a1,fc2,fcu,fg,lv,na 30    6731.0  52.53 *
         199) nxtphon:hv,li,ln,nb 35   19470.0  68.91 *
    25) nxtphon:,a2,af,fa,sv 56   38020.0  72.59
      50) wsen:md,st 39    9924.0  64.05 *
      51) wsen:fn 17   18730.0  92.18 *
  13) prephon:,fcu,hv,li,ln,lv,mv,na,nb,uv 483  377700.0  75.40
    26) psen:md 212   97050.0  66.35
      52) nxtphon:a2,fa,fc1,fc2,fc3,fcu,hv,li,lv,mv,nb,uv 126   49600.0  60.91
       104) phon:fc2,fc3,mv,na 59   13900.0  51.76 *
       105) phon:af,ln,lv 67   26410.0  68.97
         210) nxtphon:a2,fc1,fc2,fc3,fcu,lv 12     594.7  50.67 *
         211) nxtphon:fa,hv,li,mv,nb,uv 55   20920.0  72.96
           422) nxt2ph:a1,a2,af,fa,fc2,fcu,fg,hv,li,ln,lv,mv,na 50   14690.0  7
           423) nxt2ph:fo,nb,sv 5    2025.0 100.60 *
      53) nxtphon:a1,af,fg,ln,na,sv 86   38270.0  74.31
       106) pre2ph:a1,a2,af,fc2,fc3,fo,ln,mv,nb 39   12850.0  65.59 *
       107) pre2ph:fg,hv,li,lv,na,sv 47   19990.0  81.55 *
    27) psen:fn,st 271  249700.0  82.48
      54) pre2ph:,a2,af,fa,fc3,fg,fo,li,ln,na,nb,sv 163  126200.0  74.50
       108) phon:fc3,na 38   11330.0  58.00 *
       109) phon:af,fc1,fc2,ln,lv,mv 125  101400.0  79.52
         218) nxtphon:a1,a2,fa,fc3,fcu,mv,na,sv 39   21650.0  64.49
           436) stype:CVC,VC 22    6559.0  52.64 *
           437) stype:CSV,CSVC,CV,SVC,V 17    7998.0  79.82 *
         219) nxtphon:,af,fg,hv,li,ln,lv,nb 86   66950.0  86.34
           438) nxt2ph:,af,fc3,fcu,fg,fo,nb 23    7319.0  69.96 *
           439) nxt2ph:a2,fa,fc2,hv,li,ln,lv,mv,na,sv 63   51210.0  92.32
             878) pre2ph:,a2,af,fa,fc3,fg,na,nb,sv 53   40070.0  88.13
              1756) phon:ln,mv 9    1818.0  57.78 *
              1757) phon:af,fc1,fc2,lv 44   28260.0  94.34
                3514) nxtphon:,af,fg,nb 25   15050.0  84.68 *
```

```
                3515) nxtphon:hv,li,ln,lv 19    7811.0 107.10 *
              879) pre2ph:fo,li,ln 10    5288.0 114.50 *
      55) pre2ph:a1,fc1,fc2,fcu,hv,lv,mv 108   97500.0  94.51
       110) nxtphon:a1,fc2,fc3,fcu,fo,hv,ln,na,uv 37   14650.0  72.89 *
       111) nxtphon:,af,fa,fg,li,lv,mv,nb,sv 71   56550.0 105.80
         222) nxt2ph:,a2,fc2,fc3,hv,ln,lv,nb 47   15880.0  97.91 *
         223) nxt2ph:a1,af,fa,fcu,mv,sv,uv 24   32080.0 121.20
           446) phon:af,fc1,fc3,mv,na 14   17990.0 106.10 *
           447) phon:fc2,ln,lv 10    6436.0 142.30 *
7) phon:a1,a2,fa,fg,fo,uv 147  199000.0 113.60
 14) psen:fn,md 119  118600.0 106.10
   28) phon:fa,fg,fo,uv 80   73890.0  96.49
     56) nxt2ph:af,fc2,hv,lv,mv,nb 34   32830.0  79.12
      112) pre2ph:a2,fo,sv 6    1257.0  44.33 *
      113) pre2ph:af,fc2,hv,li,ln,lv,mv,na,nb 28   22750.0  86.57
        226) prephon:fc2,lv 6    3147.0  58.17 *
        227) prephon:fc3,fcu,hv,mv 22   13450.0  94.32 *
     57) nxt2ph:,a1,fc1,fcu,fg,fo,li,ln,na 46   23230.0 109.30
      114) pre2ph:a1,af,fc2,hv,li,ln,mv 22    7669.0  98.14 *
      115) pre2ph:,fa,fo,lv,na,nb,sv 24   10280.0 119.60 *
   29) phon:a1,a2 39   21980.0 125.90
     58) pre2ph:fo,hv,lv,mv,na,sv 31   14850.0 121.10 *
     59) pre2ph:af,ln 8    3610.0 144.60 *
 15) psen:st 28   45960.0 145.10
   30) nxt2ph:fc2,li,lv,mv,na,nb 22   19290.0 130.60 *
   31) nxt2ph:fc3,fcu,hv,ln 6    5043.0 198.30 *
```