

日本語モーラタイミングについて

Mora-level effects in Japanese speech timing

ウイルヘルム N. キャンベル 匂坂 芳典
W. N. Campbell & Y. Sagisaka

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

あらまし:

モーラもしくは音節レベルのタイミング制御の存在を示す音響的な手がかりを求めて、大規模な日本語音声データベースを用いた分析を行った。音韻固有の継続時間長の特徴を除去するため、正規化した時間長伸縮尺度 (z 値) を用い、隣接音韻環境にもとづく子音、母音の平均伸縮尺度値の分散分析を行なった。この結果、モーラを構成する各音韻長は、短いモーラについては伸長し、逆に長いモーラでは短縮し、モーラ長を均一に保つ傾向が見られた。同様な伸縮傾向はモーラを構成しない隣接音韻間にもみられ、隣接するモーラ間の間隔を均一にするように働くことが判明した。

ABSTRACT:

A large database of spoken Japanese was examined for acoustic clues to the presence of a timing control at the level of the mora or syllable. Normalised measures of lengthening (z -scores) were employed to factor out the effects of any phoneme-specific durational differences, and analyses of variance were performed on mean lengthening values of both consonants and vowels factored according to previous and following contexts. A tendency to normalise the duration of the mora by lengthening segments in shorter mora and shortening segments in longer mora was found. A similar lengthening was found in segments around non-CV moraic units, that also served to normalise the distance between the surrounding morae.

1 はじめに

日本語は一般にモーラもしくは音節タイミング言語とされている。このタイミングが音響セグメントの継続時間に与える制約を調べるため、大規模な日本語音声データベースを用いて分析を行った。具体的には次の問題を検討した:

- 発音上複雑に異なる子音においても母音のオンセットを等間隔にしようとする傾向が存在するか?

モーラタイプの相違により、音韻継続時間長にはかなりの差が見られる。この差は主に子音において現れ、もしモーラ長を等間隔にしようとする傾向が実際にあるとすれば、モーラもしくは音節レベルでの階層的タイミング制御の妥当性が立証される。さらにより高度のタイミング制御を、音節もしくはモーラの枠組みで眺めることは興味深い。また日本語アクセントの研究 [2] [6] や言い誤りの分析等 [3] から音節とモーラ 2 種類の構造の違いを考える必要性が指摘されており、時間単位としての拍 (モーラ) の存在を音響信号の上から再度見直すことも有意義と思われる。このためさらに次の問題を取り上げた:

- 音声信号からモーラの調音の証拠が認められるか?
- 音素あるいは音節をスピーチタイミングにおける適切な単位とみなせるか?

これまでの時間長補償等を示した音韻時間長の研究 [5] [6] [7] では音韻長の実測値が用いられてきた。これに対して本論文では各トークンの実測値をそのまま用いず、統計的かつ標準的な長さからの伸縮値を用いる。多くのサンプルを用いた平均値により個々のサンプルに依存した統計的に無意味な変動を除くことができる。

2 正規化した時間長伸縮

音声データは日本の雑誌および新聞から音韻バランスを考慮して抽出された 503 文音声からなる。文章はプロアナウンサーによって発声され、20kHz でサンプリングされた。スペクトログラムについては専門のラベラーが音韻の境界を決定、その音韻ラベルと継続時間長を計算機で取り扱えるファイルの形式に整備してある [4]。

モーラを単位とした制御の存在を調べるため、音韻長を正規化し、音素クラスごとにデータベースのすべての音韻について、絶対ミリ秒で測られた継続時間長と正規化した継続時間長伸縮尺度値との比較を行った。

それぞれの音韻の正規化した継続時間長伸縮尺度値 (z 値) を決定するには、まず音素の種類毎に平均と標準偏差を算出し、次に個々の継続時間長より各々の平均を差し引き、その差を標準偏差で割る。統計学のほうでよく知られているように正規分布の場合、全サンプルの 68% は $-1 \leq z \leq 1$,

Table 1: F tests for strength of association:

Effects on the vowel:

effect	df1	df2	F
v2 <- c2	33	12593	101.506
v2 <- c3	34	12592	82.567
v2 <- v3	17	12609	78.291
v2 <- v1	11	12615	20.486
v2 <- v4	17	12609	8.904
v2 <- c4	34	12592	6.576
v2 <- v0	16	12610	3.291
v2 <- c1	34	12592	3.126
v2 <- c0	34	12592	2.093

Effects on the consonant:

effect	df1	df2	F
c2 <- v3	17	9977	58.195
c2 <- c3	34	9960	40.016
c2 <- v2	13	9981	35.473
c2 <- v1	8	9986	16.275
c2 <- v0	16	9978	12.345
c2 <- c1	34	9960	9.721
c2 <- v4	17	9977	6.317
c2 <- c0	34	9960	4.098
c2 <- c4	34	9960	2.911

99%は $-3 \leq z \leq 3$ に納まる。このようなz値を用いることによって、平均値や分散の異なる音韻間の比較が可能となり、音韻長の絶対値に含まれる音韻固有の特性を除いた分析が行える。

3 近傍音韻における影響

音節列 $C_0V_0C_1V_1 - C_2V_2 - C_3V_3C_4V_4$ における中央に位置する子音 (C_2) および母音 (V_2) の伸縮尺度値 (z 値) をみる。z 値の分散分析を行った結果、隣接する音韻からの影響のうち、先行子音 (C_2) が同一モーラ中の母音 (V_2) に与える影響が最も大きい (分散分析: $F_{33, 12593} = 101.5$) ことが判明した。

表1. に示されるように以下、 C_3 から $V_2 \rightarrow V_3$ から V_2 の順で影響の大きいことが判った。 C_2 の音韻長が長ければ V_2 の音韻長は収縮し、 V_3 が長ければ V_2 も伸長する。詳細については次節で検討するが、各音韻長がそれぞれの差異を補償しようと働くこと、また各音韻長の平均値から予想されるものよりも規則的な時間間隔を形成しようと働くことは興味深い。

子音 (C_2) もほぼ同様に隣接音韻から影響を受けている。子音の場合、母音とは違って、同じモーラの母音 (V_2) の影響よりも後続モーラ (C_3V_3) の影響のほうが大きい。このように子音の場合に後続モーラ (C_3V_3) からの影響が大きく現れている理由は、後続母音 (V_3) がそれ自身の (先行) 子音 (C_3) から強い影響を受けやすいためであろうと推察される。以上のことから判るように、子音がタイミング制御に大きな影響を与え、母音は強い影響因子としては働かない。

またこれらの結果から判るように、後続する音韻からの影響が先行する音韻の影響より大きい。このことから音声タイミングでは先行的な制御が働いているものと思われる。

length of CV moraic units factored by V

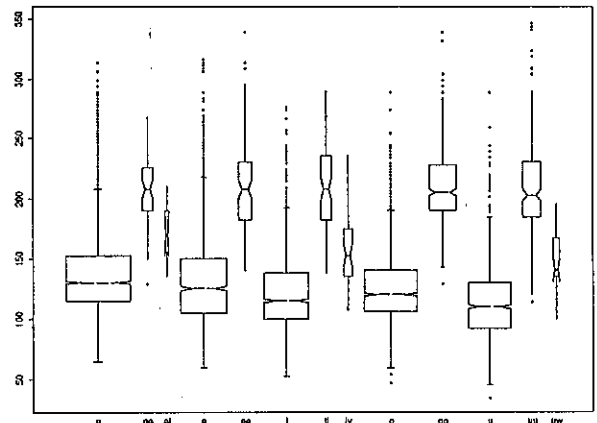


Figure 1: Mora durations factored by vowel

lengthening of CV units factored by V

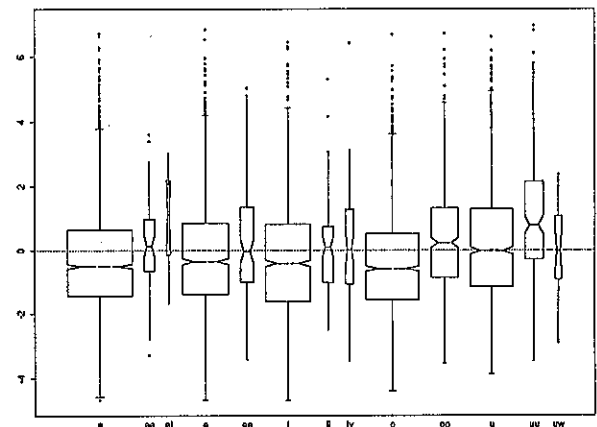


Figure 2: Mora z-scores factored by vowel

4 モーラレベルでの CV 継続時間長補償

先行及び後続音韻のグループ分けに基づき、隣接する母音と子音の継続長の和 (すなわち CV 長と VC 長) について分析した。CV グループニングと VC グループニングの影響を調べるため、音声データ中に存在する CV 接続と VC 接続それぞれについてデータを集計した。この過程で単独母音や撥音を含む箇所は除いた。時間長伸縮尺度については C, V それぞれの z 値を加算する。

図1は CV モーラを構成する母音別に測定したモーラ長 (ミリ秒) を示す。図2は母音別のモーラ長伸縮尺度値 (C, V 両音韻の z 値の合計) を示す。なお母音別に CV グループニングの子音の伸長度を見る場合、各々の z 値平均は当然ゼロとなる。図中のボックスにおける水平線は 25%、50%、75% パーセントイルを示す。ボックスの幅は各音韻環境におけるサンプル数に対応する。

これらの図の比較から判るように、実測値で表したモーラ時間長と z 値で表した時間長伸縮尺度は好対照をなしている。すなわち、ミリ秒測定でかなり長い時間長を示す長母音を含むモーラは、z 値で見ると音素固有の時間長差が除去さ

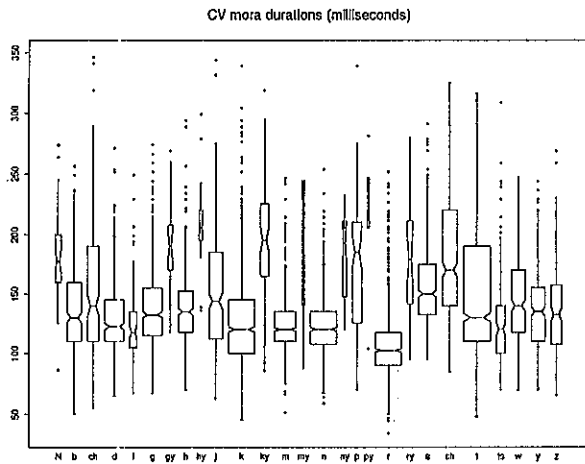


Figure 3: Mora durations factored by consonant

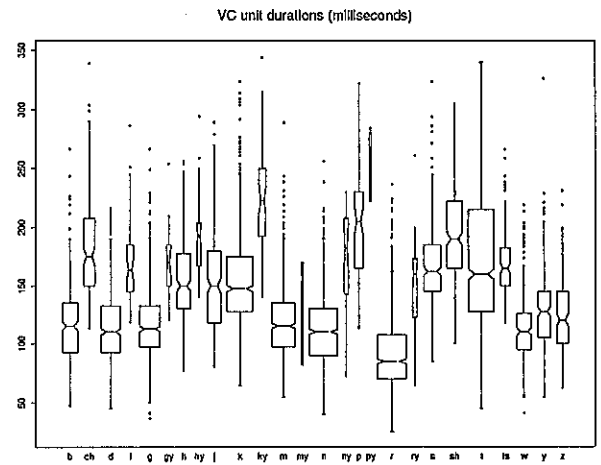


Figure 5: Vowel-Consonant (non-moraic) unit durations factored by consonant

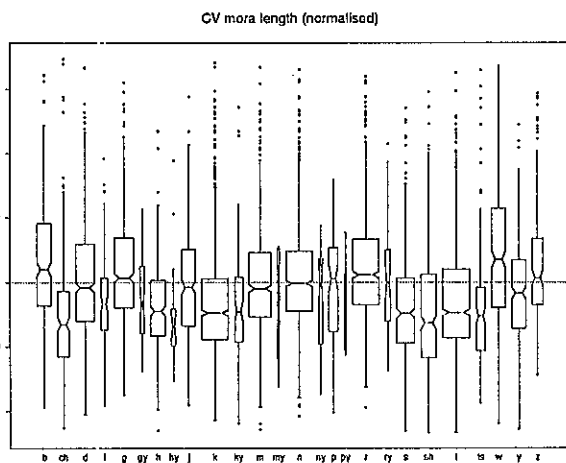


Figure 4: Mora z-scores factored by consonant

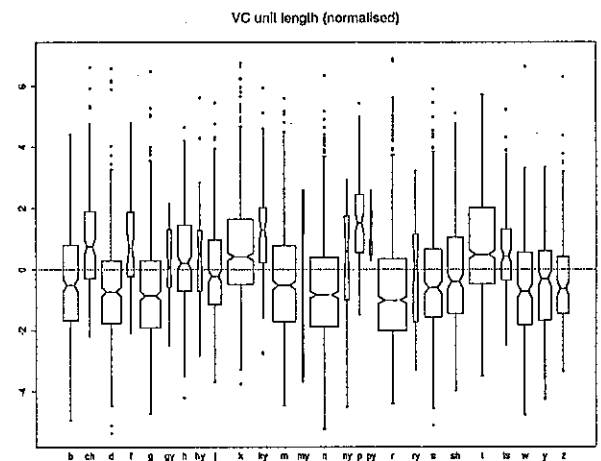


Figure 6: Vowel-Consonant (non-moraic) unit z-scores factored by consonant

れ、短母音との相違がほとんどなくなっている。また、母音別にみる時間長伸縮尺度値 (z 値) には共通した傾向がみられる。本質的に音韻長の短いもの (/u/ など) は隣接子音から伸長作用を受け、音韻長の長いもの (/a/ など) には短縮作用を受けていることが認められる。分散分析によってこの傾向は有意であることが判った。(F 12 9830 = 20.48p < 0.001) 長母音が子音に与える伸長影響については 5.2 節で検討する。

母音毎に分類して分析した図 1, 2 に対して子音毎に分類した結果を図 3, 4 に示す。これらの図から判るように、母音同様、CV 接続の継続時間長とその構成要素によって受ける短縮作用の間には負の相関関係があることがわかる。特に /ky-/、/gy-/、/t-/、/sh-/ などの子音を含む比較的長めのモーラにおいて、この傾向がはっきり見られる。これは調音上の制約によるものと思われる。逆に、短めのモーラ、たとえば /r-/、/b-/、/z-/ など子音を含む CV は伸長される。

図 5, 6 はモーラをなさない VC 接続の継続時間長と時間長伸縮尺度を示す。これらの図から相違傾向について明らかかな類似性をみることができ。VC 接続長とその伸長とは正の相関関係を示している。子音別に見た CV 接続の時間長平

均と時間長伸縮尺度値平均間の相関値は負の値、($r = -0.432$) を示し、VC 接続は ($r = 0.847$) と正の相関値を示す。両相関値が示すように、先行子音でグルーピングした場合 (モーラの CV グルーピング) のみに母音長がモーラタイミングを保とうとするよう時間長補償が行われるようにみられる。この事実は従来の実測値で示されている C-V, V-C 両方の補償傾向と若干異なっており興味深い [5] [6] [7]。

4.1 CV 単位か VC 単位か

以上の分析で最も興味深いことは VC 接続における正の相関関係と、CV 接続における、負の相関関係と言う相反する結果である。VC 接続にみられる強い正の相関関係は後に続く音韻の調音に依存するものと思われる。すなわち、長めの子音は調音に時間が必要であり、先行母音はそれぞれの器官が所定の位置に動くその間伸長維持するため、このような正の相関を生ずるのではなかろうか。

英語の場合、母音は有声破裂音の前で伸長され、無声破裂音の前で短縮される。これに対し、表 2 (before) が示すよ

Table 2: Effects of plosives on vowel length.

	unvoiced plosive:			voiced plosive:		
	mean	sd	n	mean	sd	n
before:	111.3	84.50	1754	87.7	70.66	1486
after:	73.5	31.26	2165	89.7	29.43	1667

Table 3: Lengths of segments around geminates.

	mean durations in milliseconds:					
	consonants			vowels		
	mean	sd	n	mean	sd	n
others:	53.8	26.9	8940	79.9	29.6	10132
geminate:	151.0	39.7	420	70.2	26.2	420
before:	48.3	23.4	279	91.8	19.7	290
after:	48.3	24.0	162	78.3	28.5	186

うに日本語は逆に母音は無声破裂音の前で一般に伸長され、有声では短縮される。二つの言語は音声学上異なり、単に音韻種類の差異だけでこの相違を論ずるには無理がある。CV単位を考慮することによってこの説明がつく様に思われる。日本語では母音に続く破裂音はその母音と同一の音節には属さないため、後続の子音は音節境界を越えて前の母音長を伸縮しない。これに対し、英語では多くの場合、後続の子音は同一の音節に属するため大きく伸縮される。英語でも音節境界によるブロック現象は観察されている [1]。

逆に破裂音が母音に先行する場合、表 2 (after) に示すように無声破裂音に後続する母音は短く、この結果からも時間長補償制御の存在がうかがえる。このように表面的には逆の伸長結果にもかかわらず、二つの言語における音声学ならびに継続時間長の制御機構は首尾一貫していると言える。

さらに図 3, 5 による CV 連続と VC 連続の継続時間長の分布の比較から判るようにモーラを形成する CV ユニット (SD = 43 ms) に比べて、VC ユニット (SD = 50 ms) にはかなり継続時間長の分散が大きい。このように CV を単位としたタイミング制御の存在を考慮することで、多くの観察事象が矛盾なく説明できる。次節では、このタイミング単位が特殊拍の制御にも現れることを示すと共に、単位が音節なのかモーラなのかを明らかにする。

5 特殊拍

5.1 重子音 (促音)

5.1.1 子音の継続時間長における影響:

表 3 に示すように重子音においてはかなりの継続時間伸長が明白であり、単一の子音の場合の約 3 倍の継続時間長を示す。類似の時間長補償により重子音の前のモーラ ($t_{9100} = 2.57$) と後のモーラ ($t_{9217} = 3.41$) の子音継続時間長にはある程度の短縮がみられる。これらの傾向は継続時間長パーセントイルに明確に示され、子音は少なくとも標準モーラ長になるように伸長する。この事実はタイミングと言う視点において、促音を独立したモーラとしてみなすことが妥当であることを裏付けている。

5.1.2 母音長における影響:

重子音のすぐ前にくるモーラの母音長はかなり伸長され ($t_{10420} = 6.84$)、重子音を含むモーラの母音長は短縮される

Table 4:

Vowel durations:									
Doubled vowels:			V-mora vowels:			CV-mora vowels:			
mean	sd	n	mean	sd	n	mean	sd	sd	n
123.3	40.6	1280	93.5	29.1	1609	82.8	32.6	11695	
Consonant z-scores:									
preceeding V-mora:			following V-mora:						
mean	sd	n	mean	sd	n	mean	sd	n	n
0.187	1.01	2093	0.278	0.97	2081				

($t_{10550} = 6.61$)。一方、その後続くモーラの母音には影響が見られない。これらの母音長にみる影響は先に述べた破裂音の時間長制御と一致する。子音における短縮作用は階層的タイミング制御枠組みの考えと矛盾しない。

これらの傾向は複合子音 (拗音) にはみられない。拗音では重子音に対して比較的短い平均継続時間長 80 ms (SD = 32 ms, $n=557$) を示し、一つのモーラ長を形成する傾向がみられる。

5.2 長音

長音として重複する母音の場合、長音を構成しない母音の連続および CV 中の母音に分類して母音長を比較した。結果を表 4 に示す。データベースにおける長母音は音韻長中央値 (median duration) 127.5 ms (下四分位 (lower quartile): 105.5 ms, 上四分位 (upper quartile): 150 ms) であり、それに対して単母音の場合、音韻長中央値は 80 ms (下四分位: 62 ms, 上四分位: 100 ms) である。このように長母音の音韻長は単母音の 2 倍とまではならないが 50% 近く増加する。

長母音と同じモーラ中の子音の音韻長は、中央値 77 ms (下四分位値: 54ms, 上四分位値: 105 ms) であるのに対して、単母音に先行する子音長は 47.5 ms (下四分位値: 35 ms, 上四分位値: 67 ms) と短く、母音と同様な増加率となっている。

これは長音の先行子音が時間長補償により短縮されると言う 3 節で示した傾向と対照をなす。この現象は長音節を CV + V の 2 モーラと考え、2 モーラ目の前に子音のない母音の '不十分な' 音韻長を補ない、次に続く母音のタイミング間隔を保とうとする傾向として説明することができる。

個別の単独母音が続く場合の母音 (上の表に示す V モーラ母音) は CV モーラ ($t_{13302} = 12.48$) の母音の場合、よりかなりの伸長を示すがその程度は長音にみられるほど大きくはない: 中央音韻長 = 90 ms (下四分位: 75 ms, 上四分位: 110 ms) ($t_{2837} = 29.8$) ここでは時間長補償がみられ、先行および後続モーラの子音はかなり伸長している。隣接する母音についての影響はみられない。

5.3 撥音

モーラもしくは音節がタイミング構造の基本単位として扱えるか否かは興味深い問題である。この問題の手がかりはタイミング構造にみられる /n/ の影響に求められよう。/n/ は日本語かな表記では、独立した一字「ん」を持つ。音にはその音固有のタイミングスロットがあり、おおよそひとつ分のモーラに対応する継続長をとるならば、モーラが基本単位候補と考えられる。また逆に補償的短縮の証拠があれば先行モーラが /n/ を含んだタイミング単位を形成すると考えられ、音節を単位としたタイミング制御の存在が示唆される。

Table 5: Nasal segment durations.

	min	25%	50%	75%	max
N:	25.0	75.0	90.0	107.5	185.0
n:	10.0	32.5	40.0	47.5	90.0
m:	15.0	38.1	45.0	52.5	90.0

Table 6: Consonant and vowel lengthening around /n/.

mean durations in milliseconds:

	vowels			consonants		
	mean	sd	n	mean	sd	n
others:	79.7	29.6	10285	57.3	33.8	9089
before N:	82.2	17.9	409	63.1	32.3	378
after N:	80.6	32.5	334	63.9	33.9	334

Vowel durations: F (2 11025) = 1.61 (n.s.)

Consonant durations: F (2 9798) = 10.90 p < 0.001

z-scores:	vowels			consonants		
	mean	sd	n	mean	sd	n
others:	-0.0021	0.9	10285	-0.002	1.0	9089
before N:	0.122	1.0	409	0.122	1.1	378
after N:	-0.007	0.9	334	-0.135	0.9	334

Vowel z: F (2 11025) = 37.52 p < 0.001

Consonant z: F (2 9798) = 6.17 p = 0.002

本データベースには明確に測定可能な /n/ が 466 個含まれている。表 5 にみられるようにモーラのはじめにのみ表れる /m/ と /n/ は音声上関連があり、それらの継続長と比較してみると、この鼻音モーラには二倍の継続長がある。

/n/ による影響をみるため母音、子音それぞれを /n/ の直前、直後およびに /n/ に隣接しないものの 3 種類に分類し、各々の z 値の平均値と SD を求めた。結果を表 6 に示す。表にみられるようにこれらの値にはかなりの違いがみられる。音素列 $C_1V_1NC_2V_2$ / 中で /n/ の先行母音 (V_1) と先行子音 (C_1) の音韻長は伸長し、後続子音 (C_2) は短縮した。また、後続母音 (V_2) については顕著な影響がみられなかった。

母音の平均継続時間長について表 6 をみると伸長度に差異はほとんどない。しかし z 値で比較すると先行母音 V_1 で、かなりの伸長がある。この異なる結果はパーセント値によって説明することができる。すなわち、/n/ に隣接しないサンプルは多少ではあるが極端に長い母音を含み、それが平均を偏らせている。比較的短い母音 75% までの間については先行の場合明らかに長い。

この先行音韻における音韻長伸長は別の形で 4 節で述べたタイミン単位存在を示唆している。音節タイミン単位であればこの /n/ は先行 CV と共に一音節となりタイミン補償のため CV 長が短くなるはずである。上図の比較からその傾向はみられない。逆に先行 CV 長は伸長し、/n/ の不足を補っている。これはモーラリズムを維持するために働いている。

Table 7: Vowel durations around /n/.

	min	25%	50%	75%	max
others:	12.5	60.0	75.0	95.0	255.0
before N:	30.0	71.9	82.5	95.0	150.0
after N:	20.0	60.0	75.0	95.0	197.5

と思われる。以上の実験事実は日本語において、タイミン単位は音節ではなくモーラであることを裏付けている。

6 まとめ

継続時間長の測定値そのままを用いた分析では、音韻から受ける音韻長の伸長現象の一部しか知ることができない。これに対し、音韻固有の特性を除き、継続時間長の正規化を行った時間長伸縮尺度による分析を提案した。この尺度を用いることによって音声タイミンを一層掘り下げて分析することが可能となった。

従来よく行われているように、目的とする事象を考慮した単語、文章を作成し、その音声の測定データだけを用いて音韻長制御を論ずる方法はある種の危険を伴う。すなわち、音韻長のように、人間が意識的あるいは無意識的に制御できるものは発話環境の設定が重要であると思われる。朗読音声と会話音声では明かに相違がみられる [8]。本研究では実験のために人工的に作った文章音声の比較ではなく、大規模な音声データを用いて、音韻長伸縮についてモーラ単位の影響を分析した。

本論文の各分析でモーラ長を一定の長さに保とうとする傾向(補償現象)が見られ、どの結果もモーラを日本語のタイミン単位としてみなすことにより説明できることが判明した。特殊拍の分析でもこのモーラを単位とする説明が有効であり、特に撥音 /n/ の分析により、タイミン単位が音節ではなくモーラであることが明らかとなった。

このようなモデリングは音韻レベルの制御とさらに高次の制御をうまく分離することができ、数理的な制御モデル構造の上でも有用であると思われる。

謝辞

本研究の機会を与えて下さいました ATR の諸氏に深く感謝致します。また本データベース作成にあたり、丹念な仕事をしてくださった ATR のラベラーの皆様にも厚く感謝致します。

References

- [1] Campbell, W. N. & Isard, S. D. (1991) *Segment durations in a syllable frame* Journal of Phonetics #19.
- [2] Kubozono, H. (1987) *The organisation of Japanese prosody* PhD Thesis, Edinburgh University, 1987.
- [3] Kubozono, H. (1989) *The mora and syllable structure in Japanese: Evidence from speech errors*, Language & Speech 32(3), pp 249-278.
- [4] Kurematsu, A., Takeda, K., Sagisaka, Y., Katagiri, S., Kuwabara, H., & Shikano, K. (1990) *ATR Japanese Speech Database as a tool of Speech Recognition and Synthesis* Speech Communication 9, 357-363.
- [5] Higuchi, N. (1981) 日本語連続音声における単音の持続時間に関する研究, PhD Dissertation, Tokyo University.
- [6] Sagisaka, Y. (1985) 音声合成のための韻律制御の研究, PhD Dissertation, Waseda University.

- [7] Sato, H. (1987) 規則による音声合成の研究, PhD Dissertation, Hokkaido University.
- [8] Takeda, K., Sagisaka, Y., and Kuwabara, H., (1989) *On sentence-level factors governing segmental duration in Japanese*, JASA 86, pp 2081 - 2087.