

テキスト音声合成のための韻律制御の基礎実験

島田恭宏*, 高塚有史**, 塩野 充***

*岡山理科大学大学院博士課程システム科学専攻

**三菱電機コントロールソフトウェア株式会社

***岡山理科大学工学部電子工学科

(1990年9月30日 受理)

1. まえがき

音声合成の応用が進み、我々の身近に存在する機器に組み込まれてきている。以前に話題となった喋るカメラがあった。今では自動販売機やその他家電製品も喋るようになった。これは、音声による情報伝達がより人間らしい手段であるからであり、それを機械に求めた結果と言えよう。しかし、これらの装置は、決められた語彙をそのまま装置内に記録し発声させるために、数少ない記録された語彙しか発話することしかできず、応用範囲は比較的限られたものであった。

これとは別に現在注目されるのが、発音の最小単位である音素などを記録単位とし、任意の語彙、文章を発話させるテキスト音声合成である。これが可能であれば、例えば、コンピュータのヒューマンインターフェースの改善¹⁾²⁾、視覚障害者に対する音声による情報提供³⁾⁴⁾⁵⁾など様々な応用が考えられる。原理的には、入力された漢字かな混じり文を発音させるためかな文に変換し(音韻情報の獲得)、そのかなに対応する記録単位を連結、合成すればよい。しかし、これだけでは、いわゆる流暢な日本語にはなりえず、韻律情報の制御、すなわち、アクセント、イントネーションなどの処理を必要とする。各単語のアクセントは、単独で発音した場合と、複合語、あるいは句、文章となった場合とでは、連続した単語により生起、消失、移動などの変化を受ける。本論文では、これらの変化の規則化、およびその処理手法について述べる。

2. 日本語アクセントの概要

日本語でのアクセントは、モーラ (*mora*) (拍: 母音もしくは子音+母音, 撥音, 促音, 長音, 重母音, 母音+撥音を1単位としたもの) ごとに1つの値を与えられればほぼ十分であり、感覚的には高低2つのレベルがある。そして、アクセントのあるモーラの直後でレベルが高から低に下がり、これをアクセント核と呼ぶ。 n モーラの単語(文節)には、アクセント核がないか、または、1から n モーラまでの一個所にアクセント核が存在する。このアクセント核の位置を示したものをアクセント型と呼び、アクセント核のないものを

0型アクセント, k モーラの位置にアクセント核がある場合を k 型アクセントと呼ぶ。例えば, 端, 箸, 橋はそれぞれ, 2モーラ0型, 1型, 2型である。

日本語の単語(文節)のアクセントには, 大きく分類して平板式と起上式があり, 起上式は頭高型, 中高型, 尾高型の3つがある。これをアクセントの型という(図1参照)。発音表記の横線は, その語のアクセントを示している。すなわち, 横線の部分は高く発音され, 横線のない部分は低く発音される。また, 最後の部分が鉤になっている場合は, その次の音が下がる。従って, 平板式と尾高型では, 単語単独で発音した場合は, 同じアクセントのようになるが, 付属語など基本的にアクセントを持たない単語を付けたときに, 付属語の最初の音が下がるか下がらないかにかかわっている。

アクセントの型

平板式	エイクョウ (影響)
	ダイガク (大学)
起上式	
頭高型	テレビ
	アクセント
中高型	デンドウキ (電動機)
	オカヤマ (岡山)
尾高型	ヤマ (山)
	タウエ (田植え)

図1. アクセント型の例

3. 文節内アクセント結合

いくつかの単語が連なって文節を作る場合, まず1つの文節内にはアクセント核がないか, あるいは1つあると考えられる。そこで, 1文節でのアクセント結合処理を行う必要があるが, その処理の基本となるのが2つの単語間でのアクセント結合処理(結合後のアクセントの変化は, 移動, 生起, 消失の3種類)である。いま結合する2つの単語があり, 先行単語が N_1 モーラ, M_1 型アクセント, 後続単語が N_2 モーラ, M_2 型アクセントであり, アクセント結合したことにより N_c モーラ, M_c 型アクセントの単語, または文節となる場合を考える。この時, 当然 N_c は $N_c = N_1 + N_2$ となるだけであるが, M_c については文献⁶⁾に述べられている以下に示す10パターンを用いて行う。

- | | |
|---------------------|------------------------|
| (1) $M_C = 0$ | (6) $M_C = N_1 + 1$ |
| (2) $M_C = 1$ | (7) $M_C = N_C - 1$ |
| (3) $M_C = M_1 - 1$ | (8) $M_C = N_C$ |
| (4) $M_C = N_1 - 1$ | (9) $M_C = M_1$ |
| (5) $M_C = N_1$ | (10) $M_C = N_1 + M_2$ |

今後は、この式番号をアクセント結合パターンとして、単語間、または文節間のアクセント結合について説明を行う。

(3-1) 文節内の隣接单語のアクセント結合パターン

2つの単語間でのアクセント結合パターンを各単語間のモーラ数、アクセントの型、品詞、活用形、活用型によって分類したものを以下に示す。なお分類項目は、文献のアクセント辞典⁷⁾によるものである。

- [1] **名詞+名詞** 複数の名詞が結合して複合名詞をなす場合、結合後のアクセント型は、表1の規則によって決定する。
- [2] **名詞+動詞** 名詞と動詞が結合して複合動詞をなす場合、原則的には中高型となるが、本研究では、結合後のアクセント型を表2の規則で決定する。
- [3] **名詞+形容詞** 名詞と形容詞が結合して複合形容詞をなす場合、後続単語（形容詞）が平板式であれば、複合形容詞のアクセント型は平板式か中高型となり、その他はすべて中高型となる。しかし、後続単語が平板式の場合、結果が平板式か中高型か区別がつかないため、本論文では、結合後のアクセント型を表3の規則で決定する。
- [4] **名詞+助詞** 名詞に助詞が結合する場合、複合の度合いが弱く、ほとんど名詞のアクセント型となり、規則的である。普通名詞、固有名詞、数詞すべて同様である。しかし、例外的な場合が存在し、本論文では表4に示すようなグループ分けされた助詞と、名詞のアクセント的な分類より、結合後のアクセント型を決定した。ただし、†のついているグループに対しては、他のグループに属する場合も多く、はっきりとした分類が困難なため、本論文では括弧内に示したグループに入れて処理した（以後提示する表の†についても同様である）。
- [5] **名詞+助動詞** 名詞に助動詞が結合する場合、名詞と助詞の場合と同様に複合の度合いが弱く、ほとんど名詞のアクセント型となり、規則的である。しかし、例外的な場合も含まれるゆえ、本論文では、表5に示すようなグループ分けされた助動詞と、名詞のアクセント的な分類より、結合後のアクセント型を決定した。
- [6] **動詞+名詞** 動詞に名詞が結合する場合、結合後のアクセント型は表6の規則で決定する。

- [7] **動詞+動詞** 動詞に動詞が結合する場合, 先行単語が平板式ならば結合後のアクセント型は中高型となり, 起上式ならば, 結合後のアクセント型は平板式となる。しかし, 若い年齢層では中高型で発音する傾向が強くなっている。日本語アクセント辞典でもこの中高型を大幅に取り入れており, 本論文でも結合後はすべて中高型とし, 結合後のアクセント型は式(7)で決定する。
- [8] **動詞+形容詞** 動詞と形容詞が結合して複合形容詞をなす場合, 接尾辞的な形容詞が多く, 結合後は, ほとんど中高型となるため, 結合後のアクセント型は式(7)で決定する。
- [9] **動詞+助詞** この場合, 複合の度合いが弱く, ほとんどの場合, 動詞のアクセント型となり, 規則的である。しかし, 例外的な場合もあるので, 本論文では表7に示すようなグループ分けされた助詞と動詞の文法的, アクセント的な分類より結合後のアクセント型を決定した。
- [10] **動詞+助動詞** この場合, 動詞の終止形, 連体形につく場合と, その他の活用形につく場合とで異なる。終止形, 連体形につく場合は, 複合の度合いが弱くほとんど動詞のアクセント型となる。その他の活用形につく場合は, 複合の度合いが強く全体が1つの動詞のようになるが, ほとんどのものは, もとの動詞のアクセント型を変えない。しかし, 例外的な場合もあるので, 本論文では表8に示すようなグループ分けされた助詞と, 助動詞の文法的, アクセント的な分類より, 結合後のアクセント型を決定する。
- [11] **形容詞+動詞** 形容詞と動詞が結合して複合動詞を作る場合には, 形容詞の語幹に動詞が結合する場合と, 形容詞の連用形に動詞が結合する場合がある。まず前者の場合は, 原則として中高型だが, 例外的に平板式, 中高型両様の場合もある。しかし, 本論文では, すべて中高型とし, 結合後のアクセント型は, 式(7)で決定する。後者の場合は, 先行単語(形容詞)が平板式であれば, 結合後は中高型となり, アクセント型は式(7)で決定され, 先行単語が起上式の場合は, 結合後は先行単語のアクセント型となり, 式(9)で決定される。
- [12] **形容詞+形容詞** 形容詞に形容詞が結合して複合形容詞をなす場合とは, 形容詞の語幹に形容詞が結合する場合を示す。この場合, 接頭辞的な形容詞のつくことが多く, 原則としては平板式だが, 若い年齢層では中高型に発音する傾向がある。また, モーラ数の多いものは中高型になる, 本論文では, すべてを中高型としても, それほど聞き苦しくないなので, 結合後はすべてを中高型とし, アクセント型は, 式(7)で決定する。
- [13] **形容詞+助詞** 形容詞に助詞が結合する場合は, 複合の度合いが弱く, ほとんどの場合, 形容詞のアクセント型となり, 規則的である。しかし, 例外的な場合もあるので, 本論文では, 表9に示すようなグループ分けされた助詞と, 動詞の

文法的, アクセント的な分類より, 結合後のアクセント型を決定する。

[14] **形容詞+助動詞** 形容詞と助動詞が結合する場合は, 形容詞の終止形, 連体形につく場合と, 未然形などにつく場合とで異なる。終止形, 連体形につく場合は, 複合の度合いが弱く, 結合後は形容詞のアクセント型となる。一方, 未然形につく場合は, 複合の度合いが強く, 全体が1つの形容詞のようになる。しかし, 例外的な場合もあるので, 本論文では, 表10に示したグループ分けと, 形容詞の文法的, アクセント的な分類より, 結合後のアクセント型を決定する。

[15] **助詞+動詞** 例に示すような, …+助詞+動詞のような結合をする場合, まず動詞と助詞の結合を行った後に, その結果と動詞との結合を行う必要がある。このような場合に助詞と動詞として結合規則を作成しておく必要がある。

例. 持ってくる → 持つ (動詞)+て(助詞)+くる(動詞)

本論文では, …+助詞が平板式で動詞が1段活用である場合, 結合後は平板式となり, アクセント型は式(1)で決定する。…+助詞が平板式で動詞が5段活用である場合には, 結合後は中高型となりアクセント型は式(7)で決定される。また, …+助詞が起状式であれば, 結合後のアクセント型は, …+助詞と同じになり, 式(9)で決定している。しかし, 助詞とそれに続く動詞との間を文節として区切れば, この間の結合は後に述べる文節間の結合となり, この処理は必要なくなる。しかし, アクセント的には1度に発声する方が自然である。文節間結合では, アクセント句, ポーズ位置の影響を受けて不自然な発声となる可能性があるので, できるだけ1つの文節として処理する。

[16] **その他の結合パターン** これまで色々な単語の組み合わせによる結合パターンを示してきたが, これですべてというわけではない。そこで, 他の単語について述べる。形容動詞は形容詞と同様に, 副詞, 連体詞, 感動詞は名詞と同様に扱う。接続詞は独立文節であるゆえ, アクセント的に単独で発声させる。

表1. [1] 名詞+名詞での結合パターン

後続単語の条件	結合パターン
$N_2 \geq 2$ で頭高型か中高型	(10)
$N_2 \geq 2$ で平板式か尾高式	(6)
$N_2 \leq 2$, 分類1	(5)
$N_2 \leq 2$, 分類2	(1)

名詞の分類

<分類1> : てん (点), けん (権), かい (会) など

<分類2> : てき (的), せい (性), ご (語) など

表2. [2] 名詞+動詞での結合パターン

後続単語の条件	結合パターン
平板式である場合	(9)
その他	(10)

表3. [3] 名詞+形容詞での結合パターン

後続単語の条件	結合パターン
平板式である場合	(1)
その他	(10)

表4. [4] 名詞+助詞での結合パターン

アクセントの型		A	B	C	D	E [†]
平板式		(1)		(8)	(6)	
起 状 式	頭高型	(9)				(1)
	中高型					
	尾高型	(1)				

助詞の分類

< A > : か, が, から, に, は

< B > : の

< C > : かな, かね, ね

< D > : かしら, くらい, さえ, より

< E > : な (< A > と同様の発音をする場合もある)

表5. [5] 名詞+助動詞での結合パターン

アクセントの型	a	b	c	d [†]
平板式	(1)	(6)	(8)	
起状式	(9)			

助動詞の分類

< a > : だ, である

< b > : みたい, ような

< c > : らしい

< d > : だけ (< c > と同様の発音をする場合もある)

表6. [6] 動詞+名詞での結合パターン

先行単語	後続単語	結合パターン
平板式	平板式	(1)
	起立式	(10)
起立式	——	(9)

表7. [9] 動詞+助詞での結合パターン

活用形		連用形				終止形, 連体形					假定	命令
アクセントの型	活用型	F	G	H	I	J	K	L	M	N [†]	O	P
平板式	——	(1)	(7)		(1)	(1)	(8)	(6)	(5)		(5)	(7)
起 状 式	頭高型	——	(9)		(6)	(9)				(1)	(9)	(9)
	中高型	一段	(3)									
		五段	(9)									
		サ変	(2)									

助詞の分類

- <F> : て, で
- <G> : さえ, たり, ても, つつ
- <H> : に, は
- <I> : ながら
- <J> : きり, と, ほど, もの, よ
- <K> : せ, ぞ, ね
- <L> : ぐらい, さえ, まで, ゆえ
- <M> : か, かしら, から, に, より
- <N> : だけ (<J> と同様の発音をする場合もある)
- <O> : ど, ども, ば
- <P> : と, や, よ

表8. [10] 動詞+助詞での結合パターン

活用形		終止形, 連体形			未然形			連用形		
アクセントの型	活用型	e	f	g	h	i	j	k	l	m
平板式	——	(6)			(1)			(1)		
起立式	サ変	(9)	(7)	(7)	(5)	(7)	(2)	(7)		
	その他						(9)			

助動詞の分類

- <e> : そう, だ, みたい, よう, ようだ
- <f> : らしい
- <g> : まい
- <h> : させる, せる, た, られる, れる
- <i> : ない
- <j> : う, まい, ます, よう
- <k> : た
- <l> : そう, だ, たい
- <m> : ます

表9. [13] 形容詞+助詞での結合パターン

活用形		終止形, 連体形					連用	假定
アクセントの型	モーラ数	Q	R	S	T	U [†]	V	W
平板式	——	(1)	(8)	(6)	(4)	(1)	(1)	(4)
頭高型	——	(9)					(2)	(2)
中高型	3以下						(3)	(4)
	4以上							

助詞の種類

<Q> : だけ, と, な, ほど

<R> : ね

<S> : くらい, さえ, とか, ばかり, まで

<T> : か, が, かしら, けれど, よりも

<U> : だけ (<Q> と同様の発音をする場合もある)

<V> : く, さ, て, でも, も

<W> : ど, ども, ば

表10. [14] 形容詞+助動詞での結合パターン

活用形	終止形, 連体形				未然
アクセントの型	n	o	p	q [†]	r
平板式	(6)	(7)	(4)	(7)	(7)
起伏式	(9)				

助動詞の分類

<n> : そう, だ, みたい, ようだ

<o> : らしい

<p> : です

<q> : らしい (<o> と同じ発音をする場合もある)

<r> : ろう

(3-2) 3つ以上の単語のアクセント結合

原則としてアクセント結合は、左から巡回的に適用する。例えば、“食べられます”という文節を考えると、

$$\begin{aligned} \text{タ ベ ル} &+ \text{ラ レ ル} = \text{タベラレル} \\ \text{タベラレル} &+ \text{マ ス} = \text{タベラレマス} \end{aligned}$$

となる。これを、巡回適用則⁶⁾といい図2に示すアルゴリズムで処理を行う。

巡回適用則では、結合した結果を1つの単語として扱う必要があるため、結合後の文法情報(品詞, 活用形, 活用法)を決定する必要がある。基本的には、自立語+自立語の場合は、後続単語の文法情報の文法情報を受け継ぎ、自立語+付属語の場合は、先行単語の文法情報を受け継ぐ。しかし例外として、動詞+動詞の場合、結合することによって動詞をなす場合と名詞をなす場合がある。これに関しては、後続単語が連用形, 終止形であれば、結合後は動詞, 後続単語がそれ以外であれば結合後は名詞として扱う。

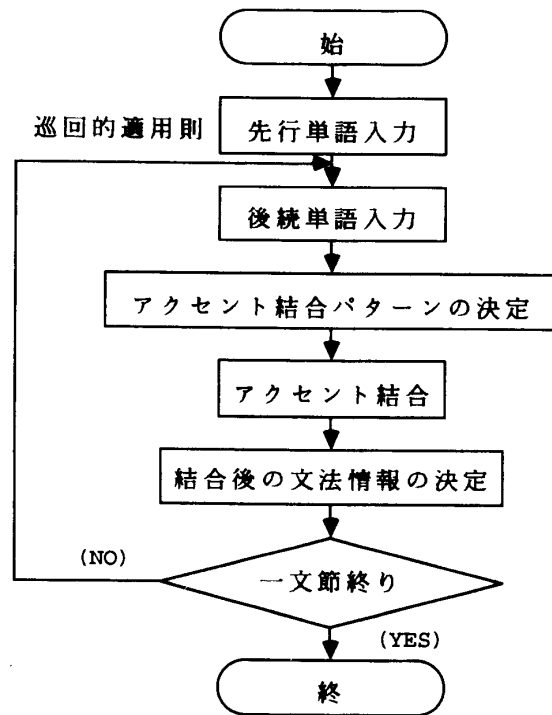


図2. 巡回的適用則のアルゴリズム

(3-3) 数詞に対するアクセント結合

文章中に数字が出現した場合、それを発声させるには表音文字（かな）に変換しておく必要がある。本論文では0～9，十，百，千，万，億，兆の基本アクセントを与えておき、これを各位毎に名詞+名詞と同様のアクセント結合処理を行っている。ただし、小数点以下については、0～9までの基本アクセントをそのまま並べて出力を行った。これらの韻律情報の制御のほか、語彙の接続などによって生じる母音の長音化などの音韻に対する処理も同時に行う。

4. 文節間アクセント結合

人間の発話過程を考えると、発話者は文が与えられたとき、聞き手に文のもつ意味、内容を正しく伝えるために、文を適当な長さを有する単位に分割し、その間に適当な長さのポーズ（息継ぎ等）を入れて発声する。この適当な長さを有する単位をアクセント句といい、一つ以上の文節からなる。このように、より自然な音声出力を得るためには、アクセント処理以外にもアクセント句、ポーズ位置の決定という重要な処理が必要となる。アクセント句の長さ、ポーズ位置、ポーズ時間を決定する主な要因として、

- (a) 文の構造上決まる文節間の結合の強さ
- (b) 文の長さ（文節境界前後の長さ）

の二つが考えられる。本論文では、(a)を文法的な文節間での係り受け、(b)をモーラ数より求め、処理を構成している。

(a)の結合の強さには、道のり方式⁸⁾を用いる。この手法は、図3に示すような係り受け構造が与えられた場合、各文節の切れ目の深さを、先行文節が受けの文節を經由して後続文節に至るまでの道のり（文節数）で表現したものであり、これを間（けん）と呼ぶ。これに対し、本論文では道のり方式を間略化し、先行文節から受けの文節に至る道のり（文節数）で切れ目の深さを表し、これを文節間距離[†]と呼ぶ。

一般に、複数の文節からなる文章を発声する場合、モーラ数、分離度が大きいほど間にポーズを入れる可能性が高く、逆に、小さいほどアクセント句として結合され可能性が高いと考えられる。そこで、モーラ数と文節間距離より2つの文節のアクセント的な分離度を式(11)で表し、表11で示されるような適当なしきい値より4つのパターンに分類し、文節間のアクセント処理を行っている。ここで、しきい値は実験によって適当と思われる値を決定した。

$$\left. \begin{aligned} \text{分離度} &= (N_1 + N_2) \times \text{文節間距離} \\ N_1 &: \text{先行文節のモーラ数} \\ N_2 &: \text{後続文節のモーラ数} \end{aligned} \right\} \quad (11)$$

ただし、副詞、連体詞が次の文節を修飾している場合、また、後続文節が副詞、連体詞を除く2モーラ以下の文節である場合は分離度=0としている。

文節間でのアクセント結合は、基本的には単語間と同様に考えることができる。したがって、2つの文節よりアクセント結合パターンを決定し、アクセント処理を行えばよいということになる。ここでは、単語間でのアクセント結合パターン番号を用いて、結合後のアクセント形を求め、その結果を表12に示す。

上述の結合パターンは、2つの文節間でのアクセント結合について述べたが、3つ以上の文節が結合する場合も存在する。基本的には3.2節（単語間の場合）と同様に、巡回適用則によって処理しているが、文節間では優先的に結合を行う場合がある。これは、後続文節が副詞、ないし、連体詞で、次の文節を修飾している場合、この部分を先に結合した後に、先行文節との結合を行う。図4に文節間アクセント結合処理のアルゴリズムを示す。

韻律情報の中には、アクセント情報の他に、イントネーション成分（話調成分、声立て成分）がある。これは、1つのアクセント句を発声するときに、声が始めは高く、次第に声門下圧などにより低くなるという成分である。しかし、これは今回使用した装置が基本的にもっている機能であり、本論文ではこのことには融れていない。

表11. 結合の強さとアクセント処理

分離度	アクセント処理
0 ~ 7	アクセント句として結合
8 ~ 24	アクセント的に切り離す
25 ~ 39	アクセント的に切り離し、短いポーズを挿入する
40 以上	アクセント的に切り離し、長いポーズを挿入する

表12. 文節間のアクセント結合パターン

先行文節	後続文節	結合パターン
平板式	平板式	(1)
	起伏式	(10)
起伏式	——	(9)

係り受け構造

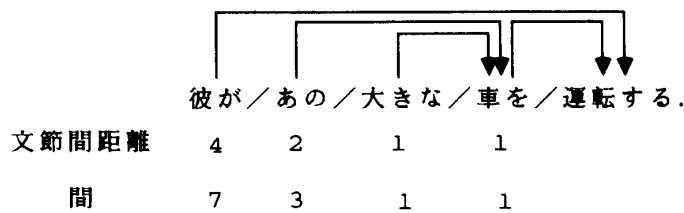


図3. 係り受け構造と文節間距離の関係

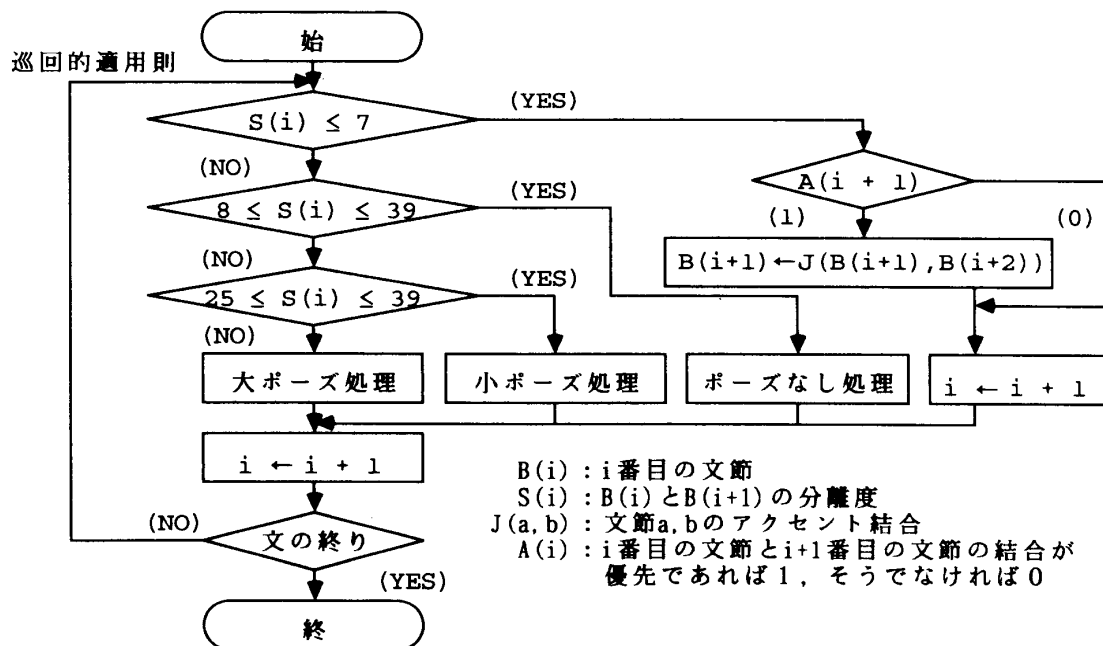


図4. 文節間アクセント処理のアルゴリズム

5. 韻律情報制御実験

(5-1) 実験結果

実験装置として今回は、パーソナルコンピュータとその内部バスに直結するタイプの規則音声合成器を用いた。入力データ、処理に必要となる辞書類は、汎用のものは作成しておらず、よって、任意の入力文に対して実験することはできない。実験を行うにあたり、まず、入力文の単語の認定、かなへの変換、および、各単語の基本アクセントを与えたデータ、係り受け構造の抽出などの処理をあらかじめ人手によって行っておき、これらのデータを用いてアクセント結合の実験を行った。以下に処理例を示しておく。

<処理例1>

原文：桃太郎。昔々あるところにお爺さんとお婆さんが住んでおりました。お爺さんは山へ芝刈りにお婆さんは川へ洗濯に行きました。

処理結果：モモ' タロー

ムカシムカシ, , ア' ルトコロニ, オジイサント オバ' アサンガ ス' ンデオリマシタ.

オジ' イサンワ, , ヤマ' エ, シバカリ' ニ, , オバ' アサンワ, カワエ センタクニ イキマ' シタ.

<処理例2>

原文：吾輩は猫である。

処理結果：ワガハイワ ネ' コデアル.

<処理例3>

原文：文字図形情報ネットワークシステムのこと。

処理結果：モジズケージョーホーネットワークシ' ステムノコト.

<処理例4>

原文：1 2 3 8 9 6 5 4。

6 5 1 3 2 4 5 6。

1 1 0 0 0 0 0 0 0。

4 4 5 4. 3 2 1 3 2 1。

処理結果：セ' ン ニヒャク' サ' ンジュー ハチマ' ン キューセ' ン

ロピャク' ゴジュ' ー ヨ' ン.

ロクセ' ン ゴヒャク' ジュ' ー サンマ' ン ニセ' ン ヨ' ンヒャク

ゴジュ' ー ロ' ク.

ジュ' ー イチオ' ク.

ヨンセ' ン ヨ' ンヒャク ゴジュ' ー ヨ' ン テン サン ニ'

イチ' サン ニ' イチ.

ただし処理結果中で「' 」はアクセント核の位置,

- 「 」は句と句の区切り,
- 「, 」はポーズと音の立て直しを伴う区切り,
- 「, , 」は長いポーズと音の立て直しを伴う長い区切り,
- 「. 」は音の下降を伴う文の終り (基本イントネーション),

を各々示す。

(5-2) 検 討

一般に、その文節がどの文節を修飾しているかは、その文節の最後の付属語によって決定される場合が多い。しかし処理例1で、“昔々”という文節は名詞が単独で1つの文節を構成しており、どの文節を修飾しているかを決定することは非常に困難である。したがってこのような場合、この文節はどの文節も修飾しないということとして処理を行った。

処理例2では、“猫である”、の部分で“…+助詞+動詞”という形になっている。これは、今回の実験で補助用言を導入していないために文節としては不適当な形となってしまう。このため本論文では、文節の形としては適当ではないが単語間の結合パターンとして“助詞+動詞”のパターンを設け対処している。これは先にも述べたが、この結合パターンが無い場合、助詞とそれに続く動詞との間を文節として区切ることとなる。しかし、アクセント的には一度に発声させた方が自然である。文節間の結合では、アクセント句、ポーズ位置の影響を受けて不自然な発声となるため、補助用言を導入していないための対処としてこの結合パターンをあえて導入している。

処理例3では、“文字図形情報ネットワークシステム”と言う長い単語に対してアクセント核が1カ所しか存在していない。これは、本実験システムが1文節でのアクセント核は1つ以下であるという考えで処理を行っており、当然の結果である。しかしながら、もっと長い単語（いくつかの短単位の単語（主に名詞）をつなぎ合わせてできる固有名詞、専門用語等の複合語）ではとても聞きづらい音声となる場合が考えられる。そこで本論文では対応していないが、単語内での各短単位の単語の関係より、本システムが文節間で行った処理と同様の処理を施し、単語内での複数アクセント句が存在するような形で対応できるのではないかと考える。ただし、これには単語の意味的な要素を理解する必要があり、構文解析を行う必要がある。

処理例4では、一つの位を一つのアクセント句として処理しているということがわかる。この手法で比較的聞きやすい音声が得られていると考えるが、“十一億”のような場合、“ジュウ”は、前後の位とまとめて発声する方が自然である。

6. むすび

日本語テキストからの音声合成における韻律情報、主にアクセントの制御について述べた。結果として示した処理例を実際に発話させ、人の耳による評価を行った結果では、比較的良好であった。しかしながら、現段階では、基本アクセント付きの単語辞書、および、

係り受け構造を抽出する構文解析部などを持たないため、任意の文章を処理することが不可能であり、評価実験を行うまでには至っていない。現在、文章を入力対象とした場合、必要となる係り受け構造の抽出における構文解析部の研究を進めており、これを、漢字かな変換における形態素解析処理⁹⁾に組み込むことにより、任意の文章にも対応可能と考える。また、基本辞書については同様に、形態素解析に用いる辞書は作成しており、この辞書の読みの項目に基本アクセントを付与することにより結合処理の前の段階までは自動的に処理が可能であると考えられる。

[脚 注]

†文献(8)の中では分離度と定義されている。しかし、本論文のアクセント的な分離の度合いを分離度と呼ぶため、ここでは文節間距離と呼ぶこととする。

参考文献

- 1) 池原 悟, 村上憲也, 宮崎正弘, 大山房史: “日本文音声出力システムの構成”, 研実報, 第35巻第2号 (1986)。
- 2) 長倉恵一, 箱田和雄, 壁谷喜義, 平原達也: “文・音声変換ユニット”, 研実報, 第37巻第4/5号 (1988)。
- 3) 小山智史, 野島秀夫, 大田 茂, 河野 勇, 城戸勝康, 長谷川貞夫: “合成音声による盲人用多目的日本語処理システム”, Human Interface, Vol. 3 (1988)。
- 4) 小山智史, 野島秀夫, 大田 茂: “盲人のためのコンピュータインターフェース～OS-TALK～”, 電子情報通信学会創立70周年総合全国大会論文集, S13-11 (1987)。
- 5) 長谷川貞夫, 小山智史, 久米 弘: “合成音声・自動点訳による盲人の新聞・大百科辞典オンラインデータベース利用”, 信学技報, ET87 (1988)。
- 6) 匂坂房典, 佐藤大和: “日本語単語連鎖のアクセント規則”, 電子情報通信学会論文誌, Vol. J66-D, No. 7 (1983)。
- 7) 日本放送協会編: “日本語発音アクセント辞典”, 日本放送出版協会 (1988)。
- 8) 箱田和雄, 佐藤大和: “文音声合成における音調規則”, 電子情報通信学会論文誌, Vol. J63-D, No. 9 (1980)。
- 9) 島田恭宏, 塩野 充: “パーソナルコンピュータを用いた小説本の自動点訳システム”, テレビジョン学会誌, [採録決定済み]

An Experiment on Prosody Control for Text-to-speech in Japanese

Yasuhiro SHIMADA[†], Yuji TAKATSUKA^{††}, Mitsuru SHIONO^{†††}

[†] *Graduate School, Okayama University of Science,
1-1 Ridaicho, Okayama, 700 Japan*

^{††} *Mitsubishi Electric Control Software Corporation*

^{†††} *Faculty of Engineering, Okayama University of Science
1-1 Ridaicho, Okayama, 700 Japan*

(Received September 30, 1990)

Following items are necessary to process for text-to-speech in Japanese.

- 1) Replacing KANJI (ideographic) characters by KANA (phonetic) letters (articulation).
- 2) Processing the prosodic information (recognition of word, BUNSETSU (word unit), accent-phrase, and so on, and extraction of position of accent kernel).

Each word has its own position of accent kernel. However, the position varies by combining some words. The details of the change of position are “movement”, “generation” and “disappearance”. Two factors decide the position of accent kernel. The first is the accent types of each word which constitutes the accent-phrase and the second is the information such as each part of speech of words being combined. In this paper, the change of position of accent kernel is made a rule by investigating these factors.