

TM-0489

韻律情報を用いた音声会話文の
文構造推定方式

大平栄二, 小松昭男, 市川 薫

April, 1988

©1988, ICOT

ICOT

Mita Kokusai Bldg. 21F
4-28 Mita 1-Chome
Minato-ku Tokyo 108 Japan

(03) 456-3191~5
Telex ICOT J32964

Institute for New Generation Computer Technology

韻律情報を用いた音声会話文の 文構造推定方式

† † †
非会員 大平榮二 正員 小松昭男 正員 市川 薫

Structure Inference Algorithm of Conversational Speech Sentence
using Prosodic Information

† † †
Eiji OOHIRA, Akio KOMATSU and Akira ICHIKAWA
Non member Member Member

† 日立製作所中央研究所、東京都
Central Research Laboratory, Hitachi Ltd., Tokyo, Japan

謝辞 : あり なし
著者分冊指定 : ① B C D

連絡先 : 大平榮二 日立製作所中央研究所 第6部
〒185 国分寺市東恋ヶ窪1-280
TEL 0423-23-1111 内(3611)

謝辞 本研究は、第5世代コンピュータ・プロジェクトの一環として、新世代コンピュータ技術開発機構ICOTからの委託により行なったものである。

あ ら ま し 韻律情報を利用して、自然な音声会話文の文構造を推定することが可能であることを示した。自然な音声会話文は発音が曖昧となるため、音韻情報のみを用いた場合、音韻や単語の正確な認識やセグメンテーションが困難となる。音声には、音韻情報の他に韻律情報がある。この韻律情報は、音声の構文境界や係り受け構造などの統語情報と関係することが明らかにされている。このため、韻律情報から、音声会話文の文構造を求められれば、会話文を理解する上で核となる単語の存在位置を限定でき、理解性能の向上が図れるものと期待される。ここに提案する方式は、次の処理から成る。1) ボーズ及び基本周波数パタンの特徴から会話文を文やアクセント節境界に分割する。2) 分割位置における基本周波数のフレーズ成分の状態を推定することにより、会話文の構文木を生成する。なお、基本周波数パタンの特徴解析にパタンマッチング的手法を用いている点も、本方式の特徴である。

外線電話取扱い業務をタスクとした平叙文の音声会話文を用いた評価実験の結果、自然な音声会話文の文構造が推定可能なことを確認した。

1. ま え が き

計算機の高度利用の進展にともない、人間と計算機との自然な対話機能の実現が望まれている。人間にとって、最も自然な情報伝達手段は音声であり、そこでは音声による会話の利用は不可欠であると考えられる。

現在、音韻情報を用いた連続音声認識の研究が多く、の研究機関で進められている。¹これらの研究においては文法的に整った文型と、丁寧な発声という制約を設けているものの、連続音声の場合、発音があいまいになるため、音韻や単語の正確な認識やセグメンテーションが困難である。このため、音韻や単語ラティスを生成し、構文や意味を用て認識候補の絞り込みが行なわれている。このような文法的に整った文型による丁寧な発声という制約は、利用者の思考の流れを中断し、音声会話の利点を著しく損なう。しかし、入力に自然な音声会話文を用いると、発音は更にあいまいになるほか、句点位置以外での発声の中断や、「えーと」などの間投詞（不要語）が出現する。したがって、自

然な音声会話文に対しては、このような音韻のみを用いたアプローチでは理解性能に限界がある。

人間は、音声を、音韻情報のみでなく、イントネーションなどの韻律情報や言語情報、知識を用いて理解している。韻律情報は、相手の会話（話し）を理解する上で重要な役割を果たしており、会話文の文構造や意味、更には、話し手の意図をも表現する。このため、会話文理解においては、韻律情報の利用は不可欠であると考ええる。

音声合成の分野の研究で、韻律情報と統語情報との間に強い関係があることが明らかにされている。^{(2),(3)} また、会話文の言語の区切りを示す韻律情報としてポーズや基本周波数の谷が知られている。このような韻律情報から音声会話文を単にセグメンテーションするだけでなく、セグメンテーションされた音声区間の間の係り受け構造などの文構造が得られるならば、音声会話文を理解する上でキーとなる単語位置の限定も可能となる。

本稿では、韻律情報を用いて自然な音声会話文の文構造の推定が可能であることを示す。提案する方式では、次の2段階により音声会話文の文構造を推定する。

- 1) ポーズや基本周波数パターン（基本周波数の高低

の時間変化)の特徴から、会話文を文や重文境界位置及び、アクセント節境界位置(1個又は、1個以上の文節が1つのアクセントを構成する単位を以下アクセント節と呼ぶことにする)で分割する。^(*)

2) 基本周波数のフレーズ成分の状態を推定することにより、分割された音声間の係り受け解析や分割位置の言語境界の判定を行ない、会話文の構文木を生成する。

2. 文構造推定方式の概要

本方式の機能構成を図1に示す。(1)まず、音声会話文を標本化し、入力した後、音声パワーや基本周波数などの特徴を抽出する。(2)そして、抽出された韻律情報の特徴や、文やアクセント節のアクセント形状を折線で近似した標準基本周波数形状を用いて、音声会話文を文やアクセント節位置で分割する。(3)次に、分割された位置を対象として、主に分割点前後の基本周波数パタンの特徴から、分割位置におけるフレーズ成分の状態を推定する。この推定結果に基づいて音声会話文の構文木(係り受け構造)を生成する。

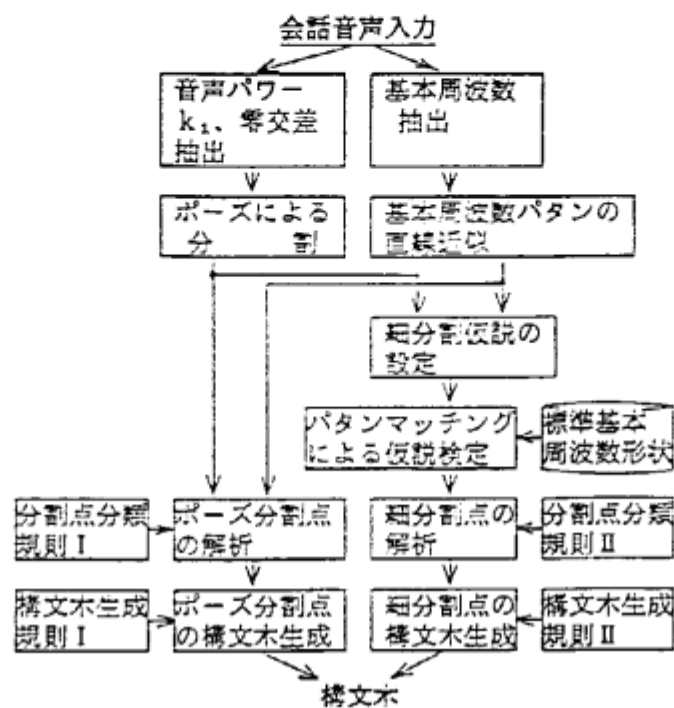


図1. 基本機能の構成
Fig1. Basic functional block diagram

1) 前処理：音声会話文は、12 KHz, 12 bitで標本化した後、7.5 msec (フレーム) 毎に、音声パワー、 k_1 (1次PARCOR係数)、零交差回数及び基本周波数を抽出する。基本周波数は、自己相関法により求め、基本周波数の連続性を利用して倍ピッチ/半ピッチの修正を行なう。そして、まず音声パワー、 k_1 、零交差回数をを用いて、各フレームを無音、有声音、摩擦音にコード化する。

2) 会話文分割：会話文を文やアクセント節位置で分割する。但し、分割された位置が文境界であるか、文節境界であるかの判定は、ここでは行なわない。本機能は、次の2つの処理からなる。①句点位置に対応する文末では、息つきなどによりポーズが生じることから、ポーズ位置を検出することにより文境界候補を求める。②我々は日常の会話において、抑揚をつけることにより相手に会話内容を理解しやすく伝えている⁽⁵⁾。このような声の抑揚は、音響的には、基本周波数パターンとして現れる。この基本周波数パターンの谷の位置で、文（重文）や文節境界が生じることが知られている。⁽⁶⁾そこで、基本周波数パターンの谷の位置を対象として、ポーズ位置で分割された音声区間の内部を、パターンマッチング的手法を用いて、更に細分割する。本処理については、3章で述べる。

3) 文構造推定：基本周波数パターンは、語義、構文に関する言語情報の伝達に重要な役割を果たしている。特に、基本周波数パターンにおけるフレーズ成分は、文構造と深い関係があり、文の統語構造からフレーズ成分を生成する規則が作られている。⁽²⁾また、分割対象としたポーズや基本周波数パターンの谷は、アクセント成分やフレーズ成分の境界において生じ、フレーズ

成分の形状はこれらの分割点に反映される。このため、分割点におけるフレーズ成分の状態を解析することにより、統語構造との対応を取ることができる。具体的には、分割点がフレーズ成分の終了位置（フレーズ成分境界位置と呼ぶ）か、フレーズ成分の下降の途中位置（同一フレーズ成分上と呼ぶ）かを推定する。本処理については、4章で述べる。

3. 会話文の分割

3.1 ポーズによる分割

会話文においては、複数の文が連続して発声される場合がある。このため、まず文境界を検出する必要がある。本処理では、息つきなどにより生じるポーズを利用して、会話文を分割する。文末のポーズ長としては、音声合成の分野の研究において約300 msec以上の報告がある。⁽¹⁾ここで、ポーズの検出においては、促音の取扱いが問題となるが、促音の持続時間は一般に、長いもので250 msec程度と言われている。⁽²⁾

しかし、次のような場合には、上記方式のみでは文境界を分割できない。文末の音韻が無声化し、長い摩擦音となる時で、このような場合には、次に続く文との間にポーズ（無音区間）がほとんど生じない場合が

ある。このような性質を示すものの例に、文末が助動詞「ます」で終了する文がある。助動詞「ます」の「す」は無声化し、長い摩擦音となる。図2は、助動詞「ます」で終了する文の無声化された「す」の摩擦音区間と、それに続く無音区間の持続時間の観測例を示したものである。本図から、両者の持続時間の和はポーズと定義した持続時間300 msec以上となっていることが判る。このため、有声音以外のフレームが300 msec以上続く区間をポーズ位置とすることとした。

ここで、ポーズは文末の他、言い淀みなどが生じた時にも生成される。このため、ポーズにより分割された位置を文境界候補と呼び、文境界か否かの判定は後述する文構造推定機能により行なう。

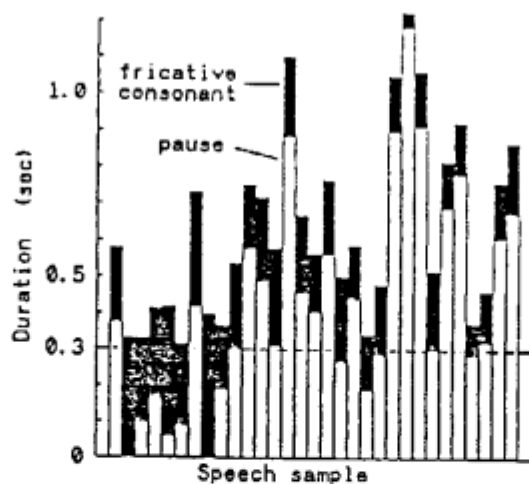


図2. 文境界におけるポーズと摩擦音の持続時間
Fig2. Duration of pause and fricative consonant at sentence boundary

3.2 基本周波数形状による細分割

本処理では、基本周波数パタンの谷の位置を対象として、ポーズ位置で分割された音声区間を、更に、文（重文）やアクセント節境界位置で細分割する。

3.2.1 基本周波数形状の直線近似

基本周波数は、無音や摩擦音においては生じないため、基本周波数パターンは歯状の形となる。このため、まず、基本周波数パタンの様子を簡潔に表現するために、文献9)と同様な方法で基本周波数の形状を折線近似する。

3.2.2 標準基本周波数形状による細分割

折線近似した基本周波数パタンの谷の位置は、文やアクセント節の境界以外に、単語内音節境界（以下単語内分割点と呼ぶ）でも生じることがある。例えば、基本周波数は摩擦音や破裂音に接続する有声音において、局所的に急峻な上昇や下降が生じる。このため、これらの位置で分割点が生じる。基本周波数パタンの谷から文やアクセント節境界のみを検出するため、以下の方法を提案する。この方法は、標準基本周波数形状とここで呼ぶ一種の標準パターンとのパターンマッチングにより、基本周波数の形状を記述し、境界を検出、分類しようというものである。

1) 標準基本周波数形状の設定

a) 形状の設定

ひとまとまりのアクセントを構成する最小単位はアクセント鉤である。このアクセント形状を直線近似すると、次の3つの形状となる。⁽¹⁰⁾

① 平板型：上昇直線＋平坦又は緩やかな減衰直線
(0型)

② 頭高型：上昇直線＋急峻な減衰直線 (1型)

③ 中高型：上昇直線＋平坦又は緩やかな減衰直線＋急峻な減衰直線 (2～N-1型)

ここで、急峻な減衰直線は、アクセント核以降のアクセントの下がり近似したもの、平坦又は緩やかな減衰直線は、アクセント上昇終了からアクセント核まで(平板型単語では語尾)のアクセント形状を近似したものである。平坦か緩やかな減衰直線かは、そのときのフレーズ成分の形状によって決まるものであるため、ここでは、両者を同一に分類し、緩やかな減衰直線と呼ぶ。このため、近似直線としては、大きく①②の三角形と③の台形のものに分けられる。

なお、平板型の文やアクセント節の場合、台形(上昇直線＋平坦＋緩やかな減衰直線)の形状を示すことがある。このため、表1のように三角形と台形の各々

の減衰直線が、フレーズの緩やかな減衰（平板型）を示す場合、又は、アクセントの急峻な減衰（起伏型）を示す場合の計4つのアクセント形状が考えられる。これを基本型と呼ぶことにする。次に、単語の一拍目の基本周波数は低く生成されることがあるため、基本型の先頭に一拍分の直線が接続される場合がある。更に、2つの文節がアクセント結合する場合、後続文節は前の文節のアクセントに低く平らに接続される。また、単語の一拍目は、二拍目と同じ高さから始まる場合もある（表1の*）。以上の検討結果から表1に示す18形状を標準基本周波数形状とした。

b) 形状の制約

標準基本周波数形状の各構成直線は、直線の持続時間、周波数変位及び勾配の値により範囲に制限を付ける。ここで、基本周波数の勾配は単位時間 t_0 （1秒）当りの基本周波数の比を用いた。すなわち、直線の持続時間を Δt 、両端の基本周波数のうち高い方を F_{max} 、低い方を F_{min} とすると、勾配 fr は、次のように求める。

$$|fr| = \frac{\Delta t \cdot F_{max}}{(\Delta t - t_0) F_{max} + F_{min} \cdot t_0}$$

ここで 上昇直線の勾配 $fr = |fr|$

下降直線の勾配 $fr = -1 * |fr|$

表1 標準基本周波数形状

	基本型	1語目の音	従属文節	標準ビッチ形状	No.
起伏型	三角 	有	有		1
			無		2
		無	有		3
			無		4*
			有		5
			無		6*
	台形 	有	有		7
			無		8
		無	有		9
			無		10*
			有		11
			無		12*
平板型	三角 	有	—		13
		無	—		14
	台形 	有	—		15*
			—		16
		無	—		17
			—		18*

*は、ポーズの後のみ生成可能な形状

なお、勾配の制約条件は、文献11)の基本周波数生成モデルを用いて決定した。

2) ポーズにより区切られた音声区間内の細分割

まず、ポーズにより区切られた音声区間内を対象にして、基本周波数パタンの谷により分割された各分割点が、文やアクセント節境界位置である場合と、単語内分割点である場合を仮定した振数の分割仮説を立てる。例えば、ポーズにより区切られた、ある音声区間の始点が t_s 、終点が t_e であり、その音声区間内に

基本周波数パタンの谷により分割された1つの分割点 t_1 があるとする。この場合、i) 分割点 t_1 を単語内分割点と仮定した場合と、ii) 分割点 t_1 を文やアクセント節境界位置と仮定した場合の下記の2つの分割仮説を生成する。下記に分割仮説において、[] の中で示される区間を文／アクセント節候補区間と呼ぶことにする。

i) [始点 t_s , 終点 t_e]

ii) [始点 t_s , 終点 t_1] + [始点 t_1 , 終点 t_e]

そして、次の処理により検定を行なう。

① 各文／アクセント節候補区間について、その範囲内の近似直線と18種の標準基本周波数形状とのマッチングを行なう。そして、最も良くマッチングした標準基本周波数形状を、その候補区間のアクセント形状とする。ここで、距離としては、近似直線と基本周波数形状との誤差（面積：Hz・sec）を用いる。

② 生成された分割仮説の内、仮説内のいずれかの候補区間が、どの標準基本周波数形状にもマッチングしなかったものは、不正な仮説として削除する。

③ 採用された分割仮説の内、距離が最も小さい（各候補区間の距離の合計）分割仮説により、会話を細分割する。

4. 文構造推定

フレーズ成分は、文構造と深い関係がある。また、フレーズ成分の形状は、ポーズや細分割点に反映される。このため、まず、ポーズ分割点および細分割点におけるフレーズ成分の状態を推定する（分割点の解析）。そして、この結果に基づいて会話文の構文木を生成する（構文木の生成）。

4. 1 ポーズ分割点の解析

ポーズによる分割点（文境界候補位置）が文境界かアクセント節境界かを判定するため、まず、その分割位置のフレーズ成分の状態を推定する。ポーズは一般的に、句点位置すなわち文境界の区切りを示すために生成される。しかし、自然な会話文の場合、ポーズは文境界以外においても生じる。例えば、言い間違い、言い淀み、又は、思考の中断が生じた時である。収録した音声サンプルにおいても「タクシー＊乗りに」の＊の位置のような複合語内部でも思考の中断によるポーズが生じる例があった。

文境界におけるイントネーションは、平叙文の場合次のような特徴を示す。

1) 文頭において、フレーズ、アクセント成分が立ち上がるため、これに伴い基本周波数は高く立ち上が

る。

2) フレーズ成分は、緩やかに下降し、文末において基本周波数は低くなる。

表2 ポーズ分割点分類規則

無音区間長 $P\lambda$ (sec)	末尾の 基本周波数 F_e (Hz)	基本周波数の差 ΔF (Hz)	摩擦音の有無	構文境界
$P\lambda > 0.3$	$F_e \leq \theta F_e$	$\Delta F \geq \theta_1$	—	フレーズ境 界
		$\Delta F \geq \theta_2$	有	同 一
		$\Delta F < \theta_2$	有	同 一
		$\Delta F < \theta_1$	無	フレーズ上
	$F_e > \theta F_e$	$\Delta F \geq \theta_1$	有	境 界
		$\Delta F < \theta_2$	有	同 一
		$\Delta F < \theta_1$	無	フレーズ上

参考図

次に、思考の中断などにより生じるポーズ位置における基本周波数パターンについて調べた結果、フレーズ成分はポーズ位置で下降を一時停止し、ポーズ後の音声において下降を再開する傾向があることが判った。すなわち、文境界以外で生じるポーズ位置の場合には、

- 1) ポーズ前後の基本周波数の差は小さく、
- 2) ポーズ前の語尾の基本周波数は高いまま保持される。

このため、表2に示す規則（分割点分類規則1）により、ポーズ位置がフレーズ成分境界位置か、同一フ

フレーズ成分上かを判定する。すなわち、ポーズ前の終端の基本周波数値 F_e とポーズ前後の基本周波数の差 ΔF を用いて、

1) F_e が低く、 ΔF が閾値 θ_1 (θ_1 は定数) より大きい時はフレーズ成分境界位置

2) それ以外は同一フレーズ成分上とする。

但し、文末が「ます」などの終助詞の時は、本来、基本周波数が更に下降し続ける文末が無声化するため、 F_e が高いまま終わる場合が多い。このため、

3) 文末が摩擦音区間のときは、 ΔF が閾値 θ_2 より大きければフレーズ成分境界位置とする。ここで、閾値 θ_1 、閾値 θ_2 は定数で、閾値 $\theta_1 > \text{閾値 } \theta_2$ である。

また、「えーと」などの間投詞は低く発声されることから、

4) ポーズで切り出された音声区間全体の基本周波数が低ければ (F_e の閾値以下)、間投詞として上記 1～3) の対象から外す。

ここで、 F_e の閾値は、話者毎に設定する。会話音声における基本周波数のヒストグラムから、話者の最低基本周波数 F_{min} を求める。これに文末の基本周波数の発声変動 θ (θ は定数) を加えた θF_e (F

$\min + \theta,)$ を F_e の高低の判定閾値とする。

4. 2 細分割点の解析

文またはアクセント節境界位置で分割された細分割位置が、フレーズ成分境界位置であるか、同一フレーズ成分上であるかを判定する。但し、細分割位置のフレーズ成分境界位置においては、後続するフレーズ成分が新規生成のものと、追加生成のものとがある。⁽²⁾ この両者の分類は、後述する構造木の生成において必要となるため、ここでは両者の分類も行なう。以下、同一フレーズ成分上の細分割位置を d 、異なったフレーズ成分間の境界を P_i とする。 P_i は、後続するフレーズ成分が新規生成の場合の細分割位置は P_1 、追加生成の場合は P_2 または、 P_3 に分類する。分類規則（分割点分類規則Ⅱ）を表3に示す。

表3 細分割点分類規則

(a) 分類規則 a

ΔF_i	ΔF_{i+1}	$f_0(i+1)$	分類
下降	上昇	高	P_2
		低	P_1
上昇	上昇	高	P_2
上昇	下降	高	P_3
下降	下降	—	d
上昇	—	低	d

(b) 分類規則 b

持続時間	ピッチ形状	分類
短	—	P_3
長	Aタイプ	P_2
	A以外	P_1

A:表1の2, 5, 6, 8, 11, 12

フレーズ成分の形状は、基本周波数の谷、即ち細分割点に反映される。このため、図3に示す細分割点前のアクセント形状の基本周波数変位 ΔF_i 、細分割点後のアクセント形状の基本周波数変位 ΔF_{i+1} 及び細分割点位置の基本周波数値 $f_0(i+1)$ をパラメータとして、細分割点のフレーズ成分の状態を推定し、分類する（細分割点分類規則 a）。しかし、対象とした細分割点のみでなく、その直前、又は直後の細分割点も新規のフレーズ成分が生成された位置である場合、 ΔF_i や ΔF_{i+1} は十分落ちきったフレーズ成分の末尾同志の変位となってしまうため、上記のパラメータではフレーズ成分の状態を推定できない。このため、細分割点の解析は、このような細分割点を対象とした分類規則（細分割点分類規則 b）を加えた、表3に示す2つの分類規則を用いて行なう。

（a）細分割点分類規則 a

分類規則を表3（a）に示し、以下に説明する。

①分割点が同一フレーズ成分上にある場合、前後の基本周波数形状は共に緩やかに減衰するフレーズ成分上にある。このため、 ΔF_i 、 ΔF_{i+1} が共に下降を示す場合は $f_0(i+1)$ の値に無関係に同一フレーズ成分上の分割点と見なす。この分割点を d とする。

② 分割点が新規フレーズ成分境界位置にある場合は、ポーズ分割位置と同様な特徴を示す。このため、 ΔF_i は下降を示し、 ΔF_{i+1} は上昇を示す。更に、分割点前のフレーズ成分は十分落ちきるため、 $f_0(i+1)$ は、低い値を示す。このような分割点をP1とする。

③ 分割点が追加フレーズ成分境界位置にある場合は、分割点前のフレーズ成分が十分落ちきらないうちに次のフレーズ成分が生成される。このため、 $f_0(i+1)$ が高い値を示す場合は、①の場合を除いて追加フレーズ成分境界位置と見なし、この分割点をP2とする。

但し、表3(a)において、 ΔF_i が上昇、 ΔF_{i+1} が下降を示す場合の分割点はP3とする。これは、このような特徴を示す大半の境界位置では、アクセント形状から細分割位置でフレーズ成分の追加生成があると推定できる。しかし、一部には、フレーズ成分の追加生成が在るとも無いともどちらともとれる場合がある。また、このような境界は、他の追加フレーズ成分境界位置と比べて、前後の音声の結び付きが強い構文境界で生じることから、他の追加フレーズ成分境界位置と区別する。

④ ②以外で $f_0(i+1)$ が低い値を示す場合は、文や文節の末尾における基本周波数のゆらぎによるもの

であると考えられる。このため、同一フレーズ成分上とし、この分割点を d とする。

(b) 細分割点分類規則 b

分類規則を表 3 (b) に示し、以下に説明する。

対象とした細分割点および、その直前、又は直後の細分割点が新規フレーズ成分境界位置である場合は、図 3 において次のような特徴を有する。

① 十分落ちきったフレーズ成分末尾のため、分割点の周波数の値 (f_0) が低く、② 新規フレーズ成分の

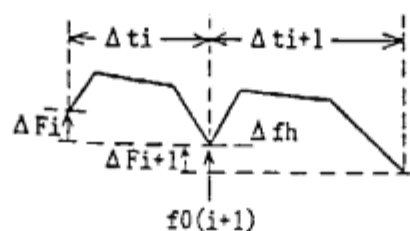


図 3 構文境界の基本周波数パターン
Fig.3 Reference patterns for F0 contours
at syntactic boundaries

生成に伴い分割点後の上昇 (Δf_h) が大きく、③更に、十分減衰するまで続くフレーズ成分であるため、持続時間 (Δt_i 又は Δt_{i+1}) が長い。このため、

- 1) 対象とした分割点で①と②の特徴を示し、
- 2) 対象とした分割点の直前、又は直後の分割点でも①と②の特徴を示すか、あるいは、そこがポーズ分

割位置のフレーズ成分境界位置であり、

3) 更に、2) の特徴を示す前又は後ろの分割点との間の持続時間 (Δt_i 又は Δt_{i+1}) が長い場合は、分類規則 a の結果とは無関係に無条件にフレーズ成分境界位置とする。

ここで、フレーズ成分の新規生成は文境界のみでなく、呼気の生理的制約により、モーラ数の多い文節境界においても生じる。このような文節の末尾の基本周波数パターンはアクセントの下降を示す。このため、対象とした分割点の前の基本周波数形状がアクセント下降を示す場合は、追加フレーズ成分境界位置とし、それ以外の形状の場合は、新規フレーズ成分境界位置

(P1) とする。ただし、追加フレーズ成分境界位置のうち、対象とした分割点に先行する分割点のうち、上記2) の特徴を示す最初の分割点と、対象とした分割点との間の(複数の分割点を含む場合もある)持続時間が長い場合は、短い場合に比べて分割点前後の音声の結び付きが弱い。このため、持続時間が長い場合はP2、短い場合はP3に分類する。ここで、アクセント下降を示す基本周波数形状とは、表1の2、5、6、8、11、12(タイプAと呼ぶ)の形状である。

ここで、 f_0 の高低の判定閾値、 Δf_h の大小の判定

閾値は、ポーズ分割位置の F_e 及び、 ΔF と同じ判定閾値を用いた。

4.3 構文木の生成

ポーズ位置および、細分割位置の分割点の解析結果に基づいて、次の規則により構文木を生成する。

1) ポーズ位置の構文木生成規則

ポーズ位置が同一フレーズ成分上にある場合は、アクセント節境界とする。次に、フレーズ成分境界位置にある場合は、文／重文境界とする。但し、文末が摩擦音区間の時は、終助詞「ます」などで終わる文境界と考えられるため、文境界とする(構文木生成規則Ⅰ)。

2) 細分割位置の構文木生成規則

表4において、まず、 $P1$ は重文境界、 d 、 $P3$ 、 $P2$ はアクセント節境界とする。そして、 $P1-P1$ 間は単独の構文木に、 d は左枝分れ構造、 $P2$ 、 $P3$ は埋込み構造とし、 d 、 $P3$ 、 $P2$ の順に構文木を生成する(構文木生成規則Ⅱ)。

以上の推定結果の例を図4に示す。

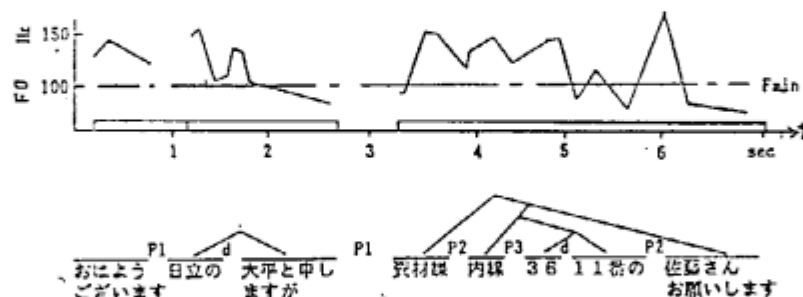


図4 文構造推定結果の例

Fig4. An illustrative example of experimental results from structural analysis

5. 実験結果

外線電話の取扱い業務をタスクとして、平叙文から成る成人男性2名の40会話文を用いて実験を行なった。音声の収録に際しては、できるだけ自然な発声を得るため、「おはようございます。〇〇でございます。」というオペレータの声を提示することにより行なった。

1) 会話文分割

ポーズ位置及び基本周波数の極小位置による分割結果を表4に示す。分割対象数は、文境界（全て挨拶文「おはようございます」）22箇所、重文境界30箇所、アクセント節境界113箇所の計165箇所であった。分割結果において、1モーラ以内の誤差は正解

とした。ここで、分割対象外のアクセント結合が生じた文節境界は135箇所であった。例えば、「○○さん * お願いします。」の * 位置である。¹²

まず、文境界の22例は全てポーズ位置で分割された。また、重文境界は30例中24例がポーズ位置で、

表4 実験結果

(a) 分割結果

分割位置	構文境界		
	文	重文	文節
ポーズ	22/22	24/30	7/113
F0	0	6/30	105/113
合計	22/22	30/30	112/113

(b) 構文境界判定結果

分割位置 判定結果	ポーズ			F0	
	文	重文	文節	重文	文節
文	22	0	0	—	—
重文	0	24	6	6	8
文節	0	0	1	0	97

残り6例は基本周波数の極小位置で全て分割できた。

アクセント節境界は113例中7例がポーズ位置で分割された。この7例中6例は、「内線～の」のような長い連体修飾(12～15モーラ)の終端である。次に、基本周波数の極小位置での細分割では、残り106例の内、「私 * 中研の～」の * 位置の1文節のみ分割できなかった。この * 位置では、短い無音区間(200ms)が存在し、かつ、無音区間前後のピッチ周波数の差がほとんど無かったためである。

また、湧き出しによるミスは6例有り、内4例は単語内分割点で分割されたものであり、2例は「申し*ますけども」の*位置に副次アクセント¹³が生じた発声であった。前者の湧き出しによるミスは、アクセント形状の接続規則を設けることにより取り除くことが可能と考えられるが、後者のミスはアクセント節境界との判別が困難である。以上の結果、表4(a)に示すように会話文の分割に関しては、96%の性能が得られた。

2) 文構造推定

ポーズ位置で分割された境界では、文境界(22例)及び重文境界は全て正しく判定された。しかし、アクセント節境界は、短い連体修飾(4モーラ)の終端位置の1例がアクセント節境界と判定できたのみで、長い連体修飾句の終端6例は全て文/重文境界と判定された。

次に、細分割位置では、重文境界の6例は全て正しく判定された。しかし、長い連体修飾句(12~15モーラ)の終端である「内線~の*○○さん~」の*位置が、10例中8例、P1すなわち重文境界と推定された。この結果、表4(b)に示すように、会話文の文やアクセント節境界の判定性能は87%であった。また、

構文木の係り受け関係の推定では、「6部の*○○さん～」の*位置のような、1つの文節から成る短い述体修飾は、後続する文節と一息で発声される場合があった。このため、このような文節の終端位置29例の内、9例が左枝分れ構造(d)のアクセント節境界に推定された。

6. あとがき

韻律情報を用いて自然な音声会話文の文構造が推定可能であることを示した。ここでは、ポーズや基本周波数パタンの特徴を用いた会話文の分割法と、分割点における基本周波数パタンのフレーズ成分の状態の推定に基づいた構文木生成法を提案した。

本方式より、音声会話文の文やアクセント節境界位置での分割、及び、分割された音声区間の係り受け構造の解析が可能となる。これらの情報を用いることにより、キーとなる単語位置を限定できるため、音声会話文理解の性能向上を図れると期待される。

外線電話の取扱い業務をタスクとした音声会話文を用いた評価実験の結果、次のような問題点が明らかになった。文や重文境界位置は全て分割できた。しかし、アクセント節境界の分割では、113例中1例の脱落

が生じ、また、6例の湧きだしが生じた。このうち、副次アクセントが生じた位置における湧きだしミスは、そのアクセント形状がアクセント節と同様の形状を示すため、取り除くことは困難である。次に、構文木生成において、モーラ数の多い連体修飾句は、フレーズ成分の生成が文と同様であるため、このような連体修飾句末尾のアクセント節境界と文境界との判別は困難であることが判った。ただし、このようなアクセント節境界は、「内線～番の」とか「～部～課の」といった意味的なまとまりを持った連体修飾句末尾のアクセント節のみで生じる。

今回は、平叙文を対象として検討を行なったが、疑問文、強調のある文についての検討が今後の課題である。なお、韻律情報から得られる文構造が、話し手のある意図を反映したものと解釈できる場合は、必ずしも書き言葉における文構造と同じ結果である必要はないものと考えられる。例えば、発話者が一つのまとまったものと考えている場合は、それを表す複数の文節列を一つのアクセント節で発声することもありえる。このような発話者の意図と音声の文構造の関係も今後の検討課題の一つである。

文 献

- 1) 新美康永：“音声認識”、共立出版（1979-10）
- 2) 広瀬啓吉、藤崎博也、山口幹雄、横尾 真：“統語構造を利用した日本語文音声の基本周波数パターンの合成”、音声研究会資料 S 8 3 - 7 0 (S59-1)
- 3) 箱田和雄、佐藤大和：“文音声合成における音調規則”、信学論（D），j63-D pp.715-722(1980)
- 4) 藤崎博也、広瀬啓吉、高橋 登、横尾 真：“連続音声中のアクセント成分の実現”、音声研究会資料 S 8 4 - 3 6 (S59-7)
- 5) 北原義典、武田昌一、市川 薫、東倉洋一：“音声音語認知における韻律の役割”、信学論（D），j70-D pp.2095-2101(1987.11)
- 6) 箱田和雄、鹿野清宏：“韻律情報を用いたセグメンテーション・アルゴリズムの会話音声認識システムへの適用” 音学講集 1 - 7 - 1 9（昭和55-3）
- 7) 箱田和雄、佐藤大和：“文音声のポーズ挿入規則” 音声研究会資料 S 7 4 - 6 4（1975-3）
- 8) 藤崎博也、樋口宜男、星野雅孝：“促音を含む文

章の時間制御” 音学講集 1-2-8 (昭和56-10)

- 9) Akio Komatsu, Eiji Oohira, Akira Ichikawa, Hirohide Endoh: "Prosodic aids to structural analysis of conversational speech", Proc. IEEE ICASSP86 42.15(1986)
- 10) NHK : "日本語発音アクセント辞書 (37刷)"、日本放送協会
- 11) 藤崎博也 : "日本語アクセントの基本周波数パターンとその生成機構のモデル" 音響誌 27-9 pp.445-453 (1971)
- 12) 藤崎博也、広瀬啓吉、河井 恒、橋本裕士: "統語構造を利用した韻律記号導出アルゴリズム"、音学講集 2-2-12 (昭和61-3)
- 13) 匂坂芳典、佐藤大和: "付属語連鎖におけるアクセント結合の分析"、音声研究会資料 S 8 2 - 2 7 (1982-7)