

TM-0370

韻律情報を用いた
音声会話文の文構造推定

大平栄二, 小松昭男
(日立)

July, 1987

©1987, 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

大平 榮二 小松 昭男 (日立中研)

1. はじめに 音声合成の分野において、言語と韻律情報との関係について、多くの検討が行なわれている¹⁾²⁾。これらの知見を利用して、複数の文が入力される音声会話文理解における韻律情報の利用法について検討を行ない、先に、主にポーズ位置における文構造推定法について報告した³⁾。今回、ポーズ間の音声区間を対象として、基本周波数形状を利用した会話文の分割、およびフレーズ成分境界位置の推定結果に基づいた文構造推定方式について検討を行なったので報告する。

2. 文構造推定法

2.1 前処理 処理手順の概要を図1に示す。音声会話文は、12 KHz, 12 bitで標準化した後、7.5 msec (フレーム) 毎に、音声パワー、K1 (1次PARCOR係数)、零交差及び基本周波数を抽出する。そして、まず音声パワー、K1、零交差を用いて、各フレームを無音、有声音、摩擦音にコード化し、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して分割する。次に、分割された区間に対して、そのピッチ包絡を直線近似により求め、その包絡直線の極小位置を分割候補点として検出する。

2.2 標準ピッチ形状による分割 標準ピッチ形状 (ピッチ極小位置間のピッチ形状) を用いて、得られた分割候補点から、文や文節の境界を求める。ピッチ極小位置間のピッチ形状は、三角や台形型のアクセント形状に加えて、フレーズ成分上に生成される谷を生じない音韻や文節が付加される。前者は、間投詞や単語の一拍目、後者は、起伏型アクセント文節にアクセント結合した平板型アクセント文節である。図2に設定した18種類の標準ピッチ形状を示す。標準ピッチ形状の各構成直線は、その持続時間、周波数変化及び勾配の値により制限する。さて、分割処理では、まず、各候補点が分割点の時と、否の時を仮定した分割仮説を立てる。そして、仮定した分割点間のピッチパターンと標準ピッチ形状とのマッチングを行ない、各分割点間のマッチングの誤差の総和が最小となる分割仮説を解として採用することにより分割点を求める。この結果得られる分割点間の区間を1まとまりのピッチ形状と呼び、その区間のピッチ形状を、最もよくマッチングした標準ピッチ形状とする。



図1. 処理手順の概要

2.3 分割点の特徴検出 会話文の係り受け構造は、フレーズ成分に反映される。又、ピッチパターンにおける谷 (分割点) と谷の間の周波数変化はフレーズ成分の形状と関係する。そこで、分割点間の周波数変化から、分割点が同一フレーズ成分上であるか、フレーズ成分境界位置であるかを判別する。フレーズ成分境界位置の分割点は、更に、フレーズ成分が新規のものはP1に、追加のものはP2かP3に分類する。以上の推定規則を図3(a)に示す。ここで、1まとまりのピッチ形状の持続時間が長い場合、そこには単独のフレーズ成分が生成される可能性が有る。したがって、分割点間の周波数変化は、フレーズ成分の末尾同

* Structural Analysis by Prosodic Information of Conversational Speech

By Eiji Oohira, Akio Komatsu (Central Research Laboratory, Hitachi Ltd.)

士の比較となってしまう。このため、①分割点のピッチ周波数が低く、②分割点後の1まとまりのピッチ形状の上昇直線が大きな上昇を示す場合は、一意的にフレーズ成分境界位置とする。そして、分割点前のピッチ形状と持続時間によりP1~P3のいずれかに決定する(図3b)。ここで、抽出されたピッチには、生理的なゆらぎがあるため、分割点間の周波数変化を一意的に上昇、下降に決定できない。このため、ファジーのメンバーシップ関数を用いて複数のコード化を許容する。

2.4 係り受け構造推定 分割点のコード化に基づいて構文木を生成する。図3において、P1~P1間は単独の構文木に、dは左枝分れ構造とし、P2、P3は埋込み構造とする。ここで、P1は重文境界、その他は文節境界とする。推定結果の例を図4に示す。

3. 実験結果 平叙文から成る成人男性2名の40会話文(外線電話文)を用いて実験を行なった。まず、分割対象文節数は、165文節であった。但し、アクセント結合が生じた文節境界は全て対象から外した。例えば、「○○さん*お願いします。」の*位置である。このうち分割できなかったのは1文節境界のみで、「私*中研の~」の*位置であった。この*位置では、短い無音区間(200msec)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。また、湧き出しによるミスは6例有り、内4例は形状から修正可能であるが、2例は「申し*ますけども」の*位置に副次アクセントが生じた発声であった。次に、係り受け構造推定では、1つの文節から成る短い連体修飾の終端位置が、左枝分れ構造(d)又は、埋込み構造(P3)の文節境界に推定された。また、長い連体修飾(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。

4. おわりに 韻律情報を用いた音声会話文の文構造推定について検討した。この結果、アクセント結合位置を除けば96%の分割性能が得られた。また、韻律情報を用いた係り受け構造推定の見通しが得られた。今後、他のタスクの会話文による検討を進めていく。尚、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、ICOTからの委託により行なったものである。

- 参考文献
(1)広瀬他;音声研資
S83-70(S59-1)
(2)箱田他;通学論文誌
J63-D 715-722(1980)
(3)Komatsu;Proc. IEEE
ICASSP86 42.15(1986)

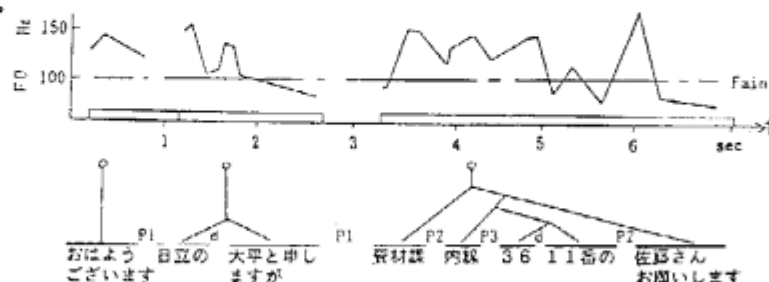
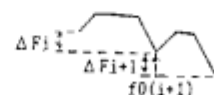


図4 係り受け構造推定例

基本型	1文節の 長	数文節の 長	標準ピッチ 形状	No.
短 形	三角	有	有	1
		有	有	2
		有	有	3
	梯形	有	有	4*
		有	有	5
		有	有	6*
長 形	三角	有	有	7
		有	有	8
		有	有	9
	梯形	有	有	10*
		有	有	11
		有	有	12*

図2 標準ピッチ形状



ΔF_i	ΔF_{i+1}	$f0(i+1)$	コード
下降	上昇	高	P2
下降	上昇	低	P1
上昇	上昇	高	P2
上昇	下降	高	P3
下降	下降	—	d
上昇	—	低	d

持続時間	ピッチ形状	コード
短	—	P3
長	Aタイプ	P2
	A以外	P1

A:図2の2, 5, 6, 8, 11, 12

図3 分割点コード化規則