

## 言語情報と非言語情報を利用した基本周波数制御の新手法

森山 高明\*      小川 均\*      天白 成一\*\*      橋本 雅行\*\*

\*立命館大学理工学部情報学科

\*\* (株) アルカディア

〒 525-77 滋賀県草津市野路東 1-1-1

〒 562 大阪府箕面市坊島 1-3-40

自然な発話の基本周波数を生成するためには、辞書化された単語のアクセント情報(言語情報)だけでなく、方言、感情、発話の個人性などを表す情報(非言語情報)が必要になると考えられる。例えば、方言には習慣化された基本周波数の変化パターンが存在する。大阪方言発話の基本周波数に見られるリズム的な要素はこの一例である。我々はこれまで、大阪方言発話の基本周波数を生成する際に非言語情報を表すパラメータとして正弦波を使用してきた。この正弦波を一意に決定するためには、周期、振幅、原点のパラメータが必要となる。このうち周期は最も重要な要素である。今までは、正弦波の周期として文節長程度の長さを与えていたが、この方法では大阪方言に見られるモーラ長レベルでの基本周波数の変化を生成することは困難である。本研究では、大阪方言の特徴が文節長やモーラ長から得られた非言語情報を利用して実現できる可能性を示す。

### A new method controlling fundamental frequency using language and nonlanguage information

Takaaki MORIYAMA\*      Hitoshi OGAWA\*

Seiichi TENPAKU\*\*      Masayuki HASHIMOTO\*\*

\*Ritsumeikan University

\*\*Arcadia, Inc.

1-1-1, Nojihigashi, Kusatsu, Shiga, 525-77 JAPAN      1-3-40, Bohnoshima, Minoh, Osaka, 562 JAPAN

It doesn't need only the language information, word accent information in dictionary, but also the nonlanguage information (i.e. dialect, emotion, and individuality) to generate fundamental frequency ( $F_0$ ) of natural utterances. For example, there are various  $F_0$  patterns like the rhythm factors in  $F_0$  of Osaka dialect. A sin wave can be used as the parameter of nonlanguage information to generate  $F_0$  in Osaka dialect. However, the following three parameters need; period, amplitude, and the origin. Especially, a period of sin wave is very important to generate various of  $F_0$  pattern. If the duration of a paragraph in sentences is used as a period of sin wave, it is too difficult to generate various of  $F_0$  pattern in Osaka dialect. In this study, a special feature of Osaka dialect can be realized using the nonlanguage information obtained from a paragraph duration or mora duration.

## 1 はじめに

人間の発話する音声は、話者の言語により大きく異なる。言語が同じであっても、方言、個人性によって

話者毎の違いがある。また、同一話者の場合でも、音声が使われる環境、発話時の感情状態などの影響で発話音声は変化する。このような音声の特徴は、主に発声、調音器官の制御方法に依存するものである。音声

の持つ情報を音韻、韻律、声質に大別すると、この制御には韻律の制御(強弱・高低・緩急・長短)が深く関わっている。韻律制御の中でも基本周波数は、単語のアクセント、文や文章のイントネーション、プロミネンスによる強調など発話様式の多様性を実現するために重要な役割を担っていると考えられる。

ところで、人間の発話音声には伝えようとする言葉の意味内容がある。また、無意味な音列であっても音声から得られる韻律情報が存在する。これらの情報は言語情報と呼ばれる。

本稿では、言語情報以外の情報を非言語情報と呼び、区別して考える。非言語情報には例えば、誰が話しているかという情報(個人性情報)や話している人の感情(情緒的情報)、方言における地方独特の言い回しなどの地域文化、プロミネンスによる強調表現などがある [1]。

したがって、言語情報だけでなく、非言語情報をも考慮した発話音声の基本周波数を生成することが可能になれば、人間にとって聞き取り易い合成音声を生成することができ、また、人間の普通の会話音声の分析にも役立つと考えられる。

そのため我々は、言語情報と非言語情報を利用した基本周波数生成手法を提案する。本手法では、言語情報は発話毎の変化や個人性に左右されない情報として、単語のアクセントを表す。非言語情報は、単語のアクセント以外の基本周波数変化に関わる情報を扱う。したがって、方言、感情、プロミネンスなどによる基本周波数への影響を非言語情報によって制御する。

言語情報と非言語情報を利用した基本周波数生成の一例として、これまでの研究では、非言語情報の中で最も扱い易いと考えられる方言、特に大阪方言が挙げられる [2]。

大阪方言発話の基本周波数には周期的な変化が見られるため、大阪方言の発話データの基本周波数に対して正弦波の非言語情報パラメータを使用することで基本周波数近似が行われている [2]。正弦波を非言語情報として使用するためには、周期、振幅、原点のパラメータを決定する必要があるが、正弦波の周期は1文節長を1周期としていた。本稿では、この正弦波の周期を変更することで、大阪方言発話の基本周波数における非言語情報の特徴を分析する。

## 2 基本周波数生成の流れ

本研究において基本周波数を生成する際には、2つの情報を必要とする(図1)。一つは言語的信息で、単語のアクセントの上昇・下降に関する情報である。もう一つは非言語的信息で、方言、感情、プロミネンスなどの影響による発話音声全体の傾向を表すパラメータがこれにあたる。

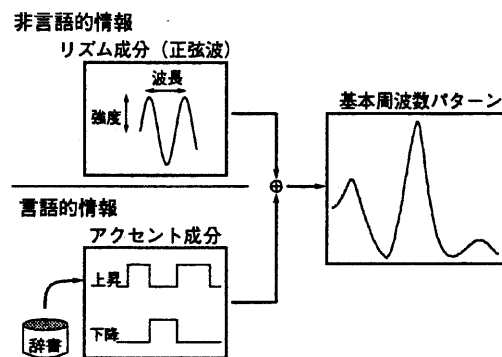


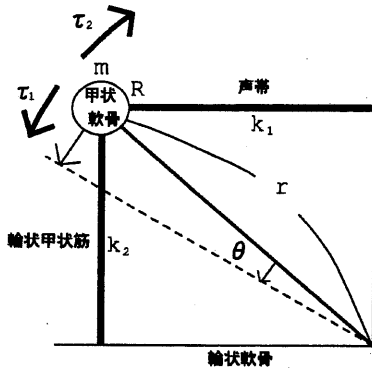
図1: 基本周波数生成の流れ

言語情報である単語のアクセント情報と、非言語情報である発話文全体の傾向から基本周波数を生成するために、基本周波数の上昇と下降を扱うことのできる基本周波数生成モデルを使用する [3]。

このモデルは、喉頭の生理機構のうち、甲状軟骨、輪状軟骨、輪状甲状筋、声帯に注目し [4]、声帯を収縮させることによって基本周波数を上げる方向に働く力( $\tau_1$ )と、声帯を伸ばすことによって基本周波数を下げる方向に働く力( $\tau_2$ )から基本周波数の変化を計算する運動力学モデルである(図2)。

## 3 非言語情報

非言語情報は、方言、感情、強調、個人性など様々な情報であり、全てを表すためのパラメータを一意に定めることは困難である。例えば、東京方言の発話文には、句頭から句末にかけて基本周波数が緩やかに下降するという傾向がある。しかし、大阪方言の発話文ではこの傾向は必ずしも見られない [5]、この例からも分かるように、東京方言と大阪方言の発話における非言語情報部分(方言の特徴)を一つの手法で表現



$k_1, k_2$ : 声帯, 輪状甲状筋のバネ定数.  
 $\tau_1, \tau_2$ : 基本周波数を上昇・下降させようとする方向に働く力.  
 $m$ : 甲状軟骨の質量.  
 $r$ : 甲状軟骨の長さ.  
 $\theta$ : 甲状軟骨の回転角.

図 2: 喉頭の運動力学モデル

することは困難であると考えられる。したがって、非言語情報を表すパラメータは、方言毎、話者毎、感情状態毎に違うものが適切に定められなければならない。先の例では、東京方言の場合は句頭から句末にかけてゆるやかに下降する直線または曲線が、非言語情報パラメータに相当する。

本稿での実験発話データは大阪方言のものであり、大阪方言発話の基本周波数には周期的な振動が見られるため、非言語情報パラメータとして、直線ではなく正弦波を使用する。正弦波を生成するためには以下の要素が必要である。

- 周期
- 振幅
- 横軸(時間軸), 縦軸(周波数軸)の原点

これらの要素のうち、大阪方言の周期的な基本周波数変化について深く考察するためには、周期の決定方法が重要になる。

## 4 発話データ

大阪方言話者男性 1 名, 女性 1 名による大阪方言発話各 70 文, 計 140 文を使用した。発話文例を図 3 に示す。

ええお嬢さんになりましたなあ  
ええなあ

図 3: 大阪方言発話文例

## 5 実験

大阪方言発話の非言語情報として正弦波を用いた実験を行っているが、これまでの研究では、正弦波の周期は実データの基本周波数中で、最初の極大から次の極大までを 1 波長としていた [2]。しかし、文節の時間長を正弦波の周期とすると、大阪方言独特のモーラ毎の急激なイントネーションの変化などを表現することは難しい。そこで本稿では、言語情報パラメータは同じで、正弦波の周期を変更して基本周波数を生成した場合、どのように周期を設定すると抽出した基本周波数とのずれが小さくなるかを調べるため実験を行う。

### 5.1 基本周波数生成方法

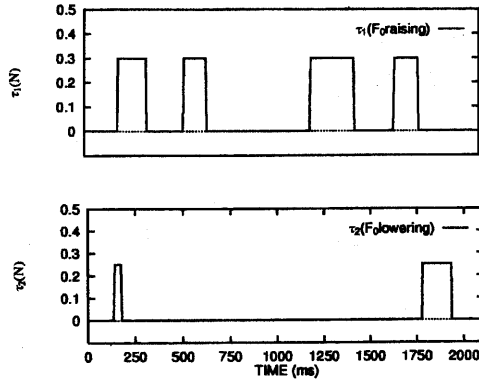
基本周波数を生成する際のパラメータを以下の方法で決定する。

#### 5.1.1 言語情報パラメータ

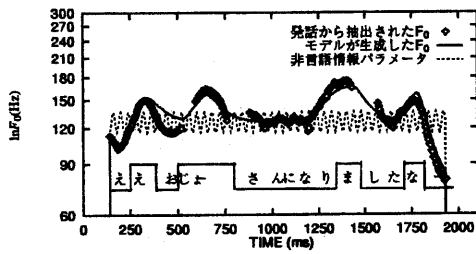
言語情報パラメータは単語のアクセントを表す。大阪方言単語のアクセントについてのアクセント辞書は存在しないため、発話データの聴取によりアクセント位置を決定する。

#### 5.1.2 非言語情報パラメータ

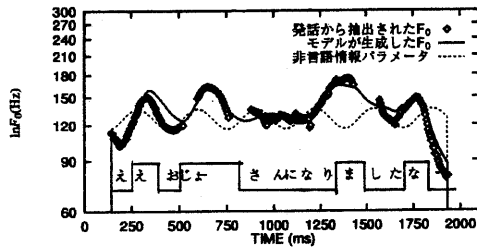
非言語情報パラメータのうち、振幅と原点は表 1 に示す方法で決定する。



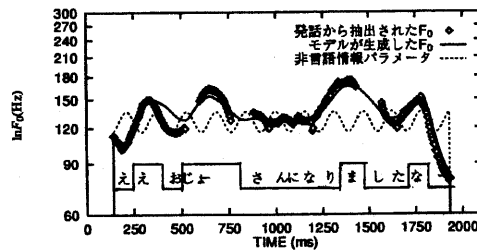
(a)  $\tau_1, \tau_2$  (言語情報パラメータ)



(b) 非言語情報パラメータ (周期: モーラ長) と  $F_0$  パターン

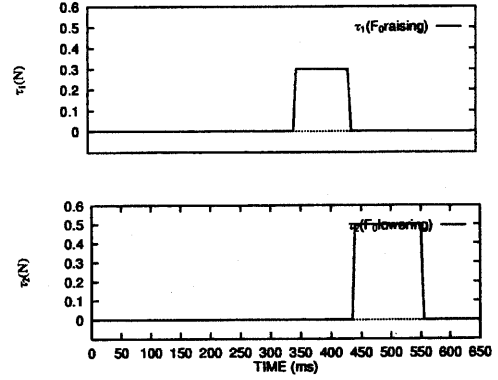


(c) 非言語情報パラメータ (周期: 文節長) と  $F_0$  パターン

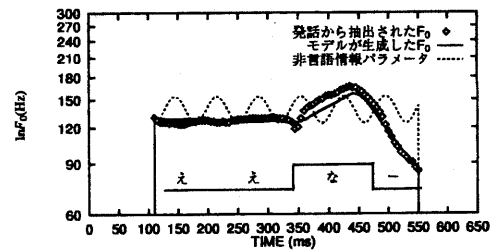


(d) 非言語情報パラメータ (周期: 連続モーラ長) と  $F_0$  パターン

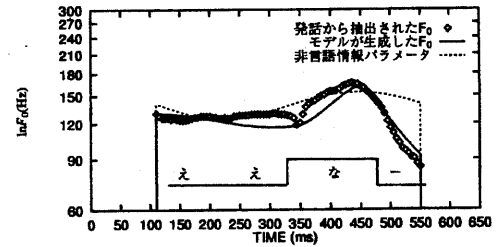
図 4: 大阪方言「ええお嬢さんになりましたなあ」



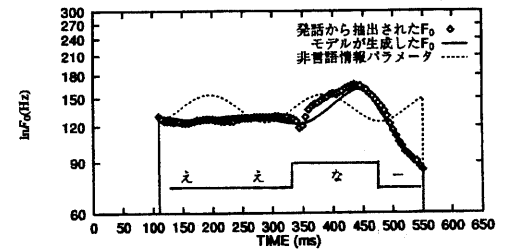
(a)  $\tau_1, \tau_2$  (言語情報パラメータ)



(b) 非言語情報パラメータ (周期: モーラ長) と  $F_0$  パターン



(c) 非言語情報パラメータ (周期: 文節長) と  $F_0$  パターン



(d) 非言語情報パラメータ (周期: 連続モーラ長) と  $F_0$  パターン

図 5: 大阪方言「ええなあ」

表 1: 振幅, 原点決定法

振幅	10~20 Hz の範囲
周波数軸	発話から抽出された基本周波数の平均値を用いる.
時間軸	発話から抽出された基本周波数の始点が正弦波の $0, \frac{\pi}{2}, -\frac{\pi}{2}$ の値

正弦波の周期は, 以下の3通りの方法で作成し, 結果を比較する.

1. 最初の1モーラ長を1周期とする.
2. 最初の1文節長を1周期とする.
3. 「ええ」などの連続したモーラは1モーラと考え(連続モーラ長), これを1周期とする.

## 5.2 基本周波数生成例

男性話者の発話文「ええお嬢さんになりましたなあ」の基本周波数生成結果を図4に示す. 図4(a)は言語情報パラメータを, 図4(b), 4(c), 4(d)中の破線は非言語情報パラメータを示す. また, 男性話者の発話文「ええなあ」の基本周波数生成結果を図5に示す. 図5(a)は言語情報パラメータを, 図5(b), 5(c), 5(d)中の破線は非言語情報パラメータを示す.

## 5.3 基本周波数のずれ

発話から抽出した基本周波数と言語情報, 非言語情報を利用して生成した基本周波数の間にはずれが生じる. 本稿では, 基本周波数を生成する際に, 正弦波の周期を2通りに変更しているため, どの周期で生成した基本周波数が実際の発話の基本周波数により近いのか定量的に調べる必要がある. 基本周波数は有声母音のみについて求められる値であるため, 当然, 比較対象となるのは実際の発話の有声母音部のみである.

ずれの計算方法を式(1)に示す.

$$diff = \frac{\sum_N \left( \frac{F_0(n) - F_{0real}(n)}{F_{0real}(n)} \right) \times 100}{N} \quad (1)$$

diff: ずれの平均 (%)

$F_{0real}$ : 発話から抽出した基本周波数 (Hz)

$F_0$ : 生成した基本周波数 (Hz)

$N$ : 有声母音部のサンプル数

## 5.4 基本周波数生成結果

図4の例では, 非言語情報パラメータを表す正弦波の周期をモーラ長にした場合と文節長にした場合, そして「ええ」のように連続して1モーラのような振舞いをするモーラ長(連続モーラ長)にした場合で比較を行った. その結果, 発話から抽出した基本周波数と生成した基本周波数のずれは表2に示す通りである.

発話から抽出した基本周波数と生成した基本周波数のずれを正弦波の周期の違いで比較した場合, 図4の例ではほぼ一致することが分かる. したがって, この点だけから見れば正弦波の周期と非言語情報との間には関係がないと言える. ただし, 図4(b)ほど短い周期での周期的振動が実際に起こっているとは考えにくいので, 単純に1モーラ長を正弦波の周期とすることには問題がある.

しかし, 図5の例では, 正弦波の周期を文節長にした場合と正弦波の周期をモーラ長にした場合で大きなずれが見られた. 正弦波の周期を文節単位にした場合, 会話文中に頻出と思われる1文節しかない文の基本周波数を生成する際に, 文節の途中から起こる基本周波数変化を自由に生成することは非常に困難である. つまり, 文節単位で正弦波の周期を決定すると大阪方言の特徴を表現しきれないことが分かる. また, 図5(b)と図5(d)を比較することで, 単純な正弦波よりも連続モーラが存在する場合は, 連続モーラ長を正弦波の周期とした方がよりずれの少ない基本周波数を生成できることが分かった.

表 2: 発話から抽出した基本周波数と生成した基本周波数のずれ

例	周期		
	モーラ長	文節長	連続モーラ長
図4	5.1%	5.1%	5.1%
図5	3.0%	5.3%	2.6%

## 6 おわりに

人間の自然な発話の基本周波数を生成するためには、単語のアクセント情報のような言語情報だけでなく、方言、感情、発話の個人性などを表す非言語情報が必要である。言語情報は、発話や話者に依存しない意味的な情報として扱われる。非言語情報は、言語情報以外の基本周波数制御に関わる全ての情報を含んでいるが、本稿では大阪方言の特徴として正弦波による周期的な基本周波数の変化を用いた。

非言語情報として正弦波を使用する際、必要となるのは以下のパラメータである。

- 周期
- 振幅
- 横軸(時間軸)、縦軸(周波数軸)の原点

特に周期は、方言のリズムなど言語情報以外の情報を表すために重要な役割を持つ。我々の研究では、正弦波の周期を発話の最初の文節長として基本周波数生成を行ってきた。しかし、この手法では、会話文中に頻出するような1文節の受け答えや、大阪方言特有のモーラ毎の基本周波数の変化を生成することは困難である。

実際の発話データから抽出した基本周波数と我々のモデルにより生成された基本周波数の有声母音部でのずれを求めた結果、図4の例では正弦波の周期を文節長とモーラ長に変えた影響は見られなかった。しかしながら、図5の例では大きなずれを生じた。

つまり、正弦波の周期を文節長にすると大阪方言の特徴であるモーラ単位での基本周波数の変化を表現することが難しくなると言える。

大阪方言発話における基本周波数は、基本的にモーラ単位の周期で周期的に変化していると考えられる。ただし、本稿で例示した「ええお嬢さんになりましたなあ」、「ええなあ」などの発話に示されるように、最初の「ええ」の部分 は、1モーラの「え」が続くのではなく、「えー」と1モーラが伸ばされた発話として扱われ、厳密に1モーラ目の「え」と2モーラ目の「え」の時間長を計ることは難しい。したがって、正弦波の周期は単純に1モーラを1周期とするよりも、

伸びた1モーラとして扱う方が発話から抽出した基本周波数に近い値を生成できると考える。

図4の例で、正弦波の周期を変更しても発話から抽出した基本周波数と生成した基本周波数のずれに差が生まれなかったのは、この連続した1モーラと文節長がほぼ一致していたからだと考えられる。

以上より、本稿で示した手法を用いることで、大阪方言の特徴を正弦波で表現できると考えられる。このように非言語情報パラメータを設定することによって、発話の非言語的な特徴を解析する手がかりになる。

今後は、より多くの発話データに対して正弦波の非言語情報を適用し、周期だけでなく、振幅、原点の定量的な決定方法を定める予定である。また、大阪方言だけでなく、他の方言、感情、個人性の特徴を解析するために本稿で示した言語情報と非言語情報の扱いが有効であることを示していきたいと考えている。

## 参考文献

- [1] 阿部匡伸: “音声合成技術がもたらすコミュニケーション革命”, *InterCommunication*, No.20 Spring, pp.166-175(1997).
- [2] 森山高明, 小川均, 天白成一: “大阪方言に見られる特徴的な基本周波数変化の分析” 音講論集 1-Q-20, pp.337-338(1997-3)
- [3] Takaaki MORIYAMA, Hitoshi OGAWA, Seiichi TENPAKU: “A new control model based on rising and falling fundamental frequency”, *Proc. of ASA and ASJ Third Joint Meeting*, pp.1171-1176(1996).
- [4] H.Fujisaki, M.Tatsumi, and N.Higuchi: “ANALYSIS OF PITCH CONTROL IN SINGING”, “*VOCAL FOLD PHYSIOLOGY*”, UNIVERSITY OF TOKYO PRESS, pp.347-363 (1981).
- [5] 杉藤美代子: “日本語アクセントの研究”, 三省堂 (1982).