論 文 の 内 容 の 要 旨

論文題目　**Quantitative Modeling, Analysis and Synthesis of Prosodic Features of Spoken Standard Chinese**
（ 標準中国語の韻律的特徴の定量的モデリングと分析と合成 ）

氏　名　倪　晋富

各言語はある特有なコードによって言語情報を伝える。 声調（tone）言語として、中国語では同一の音素列構成をもつ音節でも、声調によって別の意味を表現する。例えば、標準中国語には *ma1*「母」、*ma2*「麻」、*ma3*「馬」、*ma4*「罵」、*ma0*（疑問詞）のように四種類の声調型（Tone 1 - 4）が存在し、音声の基本周波数の動き（F0 パターン）に相当し、それぞれ **H**(igh level)、**R**(ising)、**L**(ow digging)、**F**(alling)という記号で表現される。この他にも**N**(eutral tone) という Tone 0 があり、特定的な声調型が存在しないことを示している。 これらの声調型の特徴は、個々の音節を単独に発音した場合には比較的安定であるが、多音節から構成される単語、さらに複数の単語から成る文音声の中では、個々音節の本来的声調特徴が、前後音節の声調型の影響を受けて大きく変化する。いくつかの音節が結合すると、いわゆる tone-sandhi が生ずる。 さらに、特定的な声調特徴の変化によってイントネーションが表現出来る。 これらの声調特徴の変化を正しく表現し実現することは中国語の音声情報処理で極めて重要であるが、特に中国語の音声合成において、よく知られている tone-sandhi 変化規則はこれらの現象の一部分しかカバーしておらず、 定量的記述もあまり行われていない。この観点から、本論文では標準中国語における韻律的特徴（主に F0 パターン）のモデリング、ラベリングとそれによる分析、合成を目的とする。

定量的モデルは F0 パターンの分析・合成において不可欠である。 従来から、生成過程の知見に基づいて基本周波数パターンを表現する有効なモデルが提案されているが、中国語に対しては四声による大きな起伏のため、このモデルに基づいた解析が困難であった。これに対し、本論文では基本周波数パターンの現象を統一的に表現しうる関数を提案し、それに基づいて合成に適した F0 パターンのモデルを提案している。 具体的には、正規化された周波数スケール、或は RONDO（**R**atio **O**f **N**atural frequency of driven system to natural frequency of **D**riven f**O**rce）スケールを導入することにより、 F0 パターンは山型のパターンの系列として $\Lambda(t)$ 次式で表現される。

$$\frac{\ln F_0(t) - \ln f_{0_b}}{\ln f_{0_t} - \ln f_{0_b}} = \frac{T(\Lambda(t)) - T(\lambda_b)}{T(\lambda_t) - T(\lambda_b)}, \text{for } t \geq 0,$$

where

$$T(\lambda) = \frac{1}{\sqrt{(1-(1-2\zeta^2)\lambda)^2 + 4\zeta^2(1-2\zeta^2)\lambda}}, \text{ for } \lambda \geq 1,$$

and

$$\Lambda(t) = \Lambda_{r_1}(t) + \sum_{i=1}^{n-1} \mathrm{Min}(\Lambda_{f_i}(t), \Lambda_{r_{i+1}}(t)) + \Lambda_{f_n}(t) .$$

但し、記号 *Min(y, z)*は y と z の最小のものを選択しており、f_0 と t と λ はそれぞれ基本周波数、時間、RONDO 周波数である。又、$\Lambda_{r_i}(t), \Lambda_{f_i}(t)$ は i 番目の山型の上昇と下降成分を示す。

$$\Lambda_{r_i}(t) = \begin{cases} \lambda_{p_i} + \Delta\lambda_{r_i}(1 - D_{r_i}(t_{p_i} - t)), & \text{for } t \leq t_{p_i}, \\ 0, & \text{otherwise,} \end{cases}$$

$$\Lambda_{f_i}(t) = \begin{cases} \lambda_{p_i} + \Delta\lambda_{f_i}(1 - D_{f_i}(t - t_{p_i})), & \text{for } t \geq t_{p_i}, \\ 0, & \text{otherwise,} \end{cases}$$

$$\text{where } D_{x_i}(t) = (1 + \frac{4.8t}{\Delta t_{x_i}})\, e^{-\frac{4.8t}{\Delta t_{x_i}}}, \text{for } t \geq 0.$$

ここで x は r と f を示す。このモデル基づいて、母語話者 8 名によって発声された 2509 文を分析し、モデルのパラメータの推定を行った。 実験結果から、適切なモデルパラメータを与えることにより、観測 F0 パターンにモデルを良好にあてはめることが可能だと示された。 全話者に対し、主に、ζ と λ_b, λ_t は話者や文の内容が関係なく、それぞれ 0.237、1.98、1 に固定することが可能と分かった。 f_{0_b}, f_{0_t} というパラメータは各話者の最小 F0 値と最

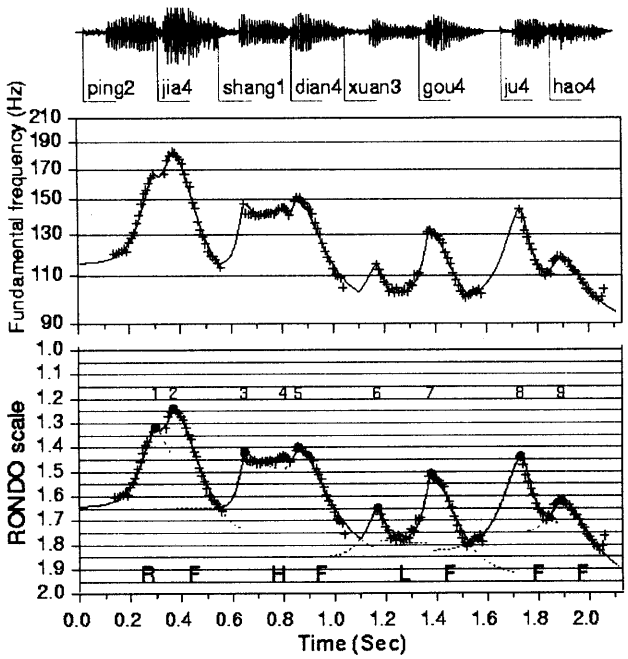

図 1．提案したモデルによる分析結果例。

表 1．図 1 の分析結果のパラメータ値

f_{0_b} (Hz)	f_{0_t} (Hz)	λ_b	λ_t	ζ	i	t_{p_i} (Sec)	λ_{p_i}	Δt_{r_i} (Sec)	$\Delta\lambda_{r_i}$	Δt_{f_i} (Sec)	$\Delta\lambda_{f_i}$
85	210	1.98	1.0	0.237	1	0.30	1.32	0.19	0.33	0.16	0.20
					2	0.37	1.24	0.08	0.11	0.32	0.54
					3	0.65	1.42	0.08	0.23	0.12	0.07
					4	0.80	1.44	0.18	0.03	0.20	0.12
					5	0.86	1.40	0.09	0.12	0.36	0.45
					6	1.17	1.65	0.16	0.20	0.08	0.15
					7	1.38	1.51	0.10	0.27	0.30	0.42
					8	1.73	1.44	0.23	0.39	0.12	0.28
					9	1.89	1.62	0.15	0.14	0.35	0.34

大F0値を示す。 又、パラメータ $\Delta t_{r_i}, \Delta\lambda_{r_i}, t_{p_i}, \lambda_{p_i}, \Delta t_{f_i}, \Delta\lambda_{f_i}$, $i=1,...,n$ の値は言語に関する情報により大きく変動する。 しかし、ある発話の F0 ピークの位置（$t_{p_i}, i=1,...,n$）を指定した場合、残りのパラメータは自動的に推定可能である。 これらのパラメータ推定を行うためのアルゴリズムを開発した。図1に一例文「平 (ping2) 価 (jia4) 商 (shang1) 店 (dian4) 選 (xuang3) 購 (gou4) 句 (ju4) 号 (hao4)」の分析結果が示されている。 図中、＋印は抽出されたF0を示し、実線と破線はそれぞれモデルによる最良近似及び山型のパターンを表す。 その分析結果に関するモデルの各パラメータの値を表1に示す。

山型のパターンにより4つの声調型は次のようなパラメトリック形式を用いて数式化される。

i番目の山型なパターンの数式化 ＜=＞ $\{\Delta t_{r_i}, \Delta\lambda_{r_i}, \Delta\lambda_{p_i}, \Delta t_{f_i}, \Delta\lambda_{f_i}\}$.

但し、$\Delta\lambda_{p_i} = \lambda_{p_i} - \hat{\lambda}_{p_i}$。 更に、$\hat{\lambda}_{p_i}$ は次の式を示すピーク・レファレンス直線によって求まる。

$$\hat{\lambda}_{p_i} = \lambda_0 + k*t_{p_i}.$$

λ_0 と k はそれぞれ直線の切片と勾配である。 基本的には、一つの山型パターンを使用してR、L、F の型を表現されて、H は2つのパターンを組み合わせて表現される。 N (Tone 0) に関しては、孤立パターンを定義する必要はない。 更に、このパラメトリック形式を用いれば、いかなる声調系列も自由に組み合わせることが出来、tone-sandhi 規則によって単語やフレーズのF0パターンが求まり、更にいくつかのイントネーション規則によって文全体のF0パターンが求まる。

Tone-sandhi 規則は、文脈による声調変化を表現するため、モデルによって定式化された。 具体的に、19 bi-、198 tri-tone-sandhi の形状をパラメトリック形式で表現した。 これらのピーク・レファレンス直線は同一で、同声調 H の系列にほぼマッチングする。これらの tone-sandhi の形状は 84 di-、538 tri-、938 tetra-syllables の単語を分析して求められたもので 19 bi-、59 tri-、221 tetra-tone-sandhi のパターンを簡潔に表すものである。 各 tone-sandhi の形状は3つのレンジ・タイプ { normal (Type A), compressed (Type B), expanded (Type C) }を有し、それによって、異なったF0パターンが実現可能である。 これらの規則の妥当性は母語話者 3 名が発声した 1730 単語と数字系列と文を用いて Analysis-by-Synthesis によって評価された。

イントネーション規則は、文全体のF0パターン生成において、談話焦点と表現意図（質問・応答など）の影響を考慮するものである。 分析データとして、朗読調と対話調で発声されたおよそ 200 文を用いた。 実験結果により、談話焦点の影響も表現意図の影響も、

tone-sandhi パターンの形状の構造を用いて適切なレンジ･タイプ選択とピーク･レファレンス直線の定義によって記述可能と示された。 談話焦点の影響は次の 3 つの条件によって記述した： pre-, under-, post-focuses。 表現意図は文末の声調のみを考慮する。 陳述文と比べて、質問文では、ピーク･レファレンス直線は平坦化され、Type B レンジ･タイプは文末の声調 **H** か **F** に適用する。 対照的に、文末の声調が **R** か **L** である場合、F0 ピークのみが上昇する。

文全体の F0 パターンの合成はテキストと発話焦点の適切な情報を用いて行った。先ず、F0 パターンのタイプを tone-sandhi 規則によって各単語毎に求める。レンジ・タイプとピーク・レファレンス直線は、各単語焦点と表現意図の条件によって選択される。 次に、声調列とレンジ・タイプを用いて選択したモデル･パラメータを、F0 ピークがピーク･レファレンス直線に一致するように調整する。 調整を行う際、ピーク・パラメータのみが再推定され、他のパラメータはそのまま残される。 最後に、これらのパラメータでモデルを制御して、特定話者の voice register へアライメントすることによって F0 パターンを生成する。この手法の実現性を調べるため、分析再合成の実験を行った。 図 2 に一つの例が示されている。 図中、＋印は抽出された F0 を示し、実線は再合成する F0 パターンを表し、直線 a、b と c はピーク･レファレンス直線である。

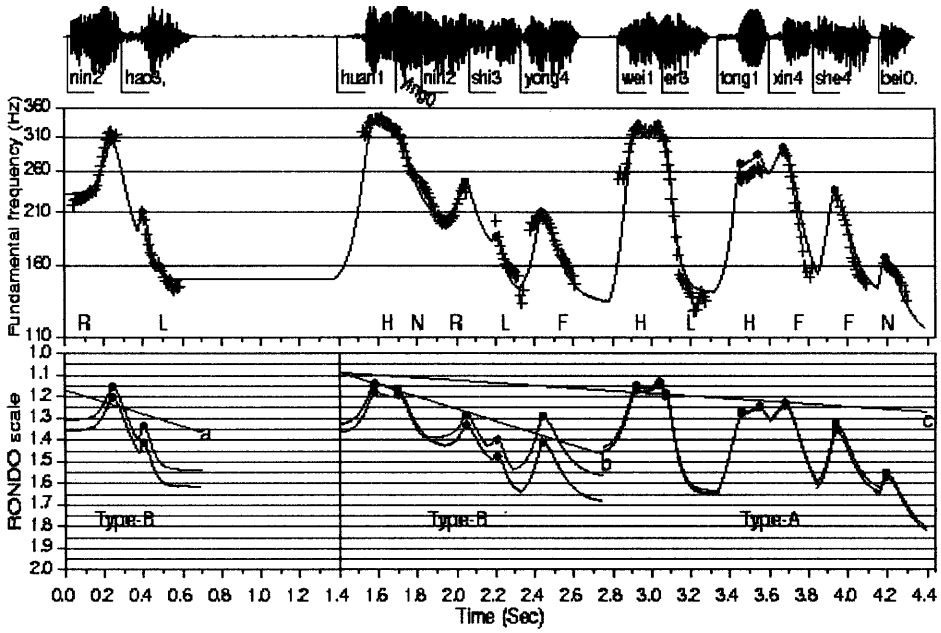

図 2．提案した手法による観測的基本周波数パターンの再合成例。

その他、 提案したモデルに基づくは観測的基本周波数パターンに対して声調の自動ラベリングも行った。 母語話者 6 名が発声した 600 文を用いて評価実験を行い。 6791 個の声調のうち、84％は正しくラベリングされた。