(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第4793776号
(P4793776)
(45)発行日　平成23年10月12日(2011.10.12)　(24)登録日　平成23年8月5日(2011.8.5)

(51)Int.Cl.　F I
G10L 13/08 (2006.01)　G10L 13/08　13OE
G10L 13/08　127C

請求項の数 3　外国語出願　(全 17 頁)

(21)出願番号　特願2005-98067(P2005-98067)
(22)出願日　平成17年3月30日(2005.3.30)
(65)公開番号　特開2006-276660(P2006-276660A)
(43)公開日　平成18年10月12日(2006.10.12)
審査請求日　平成20年2月28日(2008.2.28)

（出願人による申告）平成１６年度独立行政法人情報通信研究機構、研究テーマ「大規模コーパスベース音声対話翻訳技術の研究開発」に関する委託研究、産業活力再生特別措置法第３０条の適用を受ける特許出願

(73)特許権者　393031586
株式会社国際電気通信基礎技術研究所
京都府相楽郡精華町光台二丁目２番地２
(74)代理人　100099933
弁理士　清水　敏
(72)発明者　ジンフ・ニ
京都府相楽郡精華町光台二丁目２番地２
株式会社国際電気通信基礎技術研究所内
(72)発明者　河井　恒
京都府相楽郡精華町光台二丁目２番地２
株式会社国際電気通信基礎技術研究所内

審査官　山下　剛史

最終頁に続く

(54)【発明の名称】イントネーションの変化の特徴を声調の変形により表す方法及びそのコンピュータプログラム

(57)【特許請求の範囲】
【請求項１】
イントネーションの変化の特徴を声調の変形により表す方法であって、
話者の、個々の音節から得た語の声調の各々について、基本周波数（Ｆ０）ターゲットに関する参考値の所定の組を準備するステップを含み、前記Ｆ０ターゲットの参考値の組は、対応する語の声調を特徴づけるものであり、
前記話者のサンプル音声データ中の各音節についてＦ０ターゲット値を抽出するステップと、
前記サンプル音声データ中の各音節の前記Ｆ０ターゲット値の各々について、その音節の語の声調に関する参考値から前記Ｆ０ターゲット値への変化の度合いを表す所定の第１のパラメータを計算するステップとをさらに含み、
前記準備するステップは、
語の声調の各々について前記話者による複数個の個々の音節を録音するステップと、
それぞれの語の声調に従って、録音された個々の音節のＦ０ターゲット値を抽出するステップと、
語の声調の各々について、語の声調を特徴づけるＦ０ターゲットの各々のＦ０ターゲット値を平均して前記参考値を求めるステップとを含む、イントネーションの変化の特徴を声調の変形により表す方法。
【請求項２】
所定の第２のパラメータの分布が、当該所定の第２のパラメータの所定の基準値の両側で

つりあうように、前記所定の第１のパラメータを前記所定の第２のパラメータに正規化するステップをさらに含む、請求項１に記載の方法。
【請求項３】
コンピュータ上で実行されると、請求項１又は請求項２に記載の全てのステップを当該コンピュータに行わせる、コンピュータプログラム。

【発明の詳細な説明】
【技術分野】
【０００１】
この発明は話し言葉の処理に関し、特に、話し言葉でのイントネーションの変化を測定して所望のイントネーションの音声を合成することに関する。
【背景技術】
【０００２】
中国語の基本周波数（Ｆ０）の輪郭（一般的な意味でのイントネーション）は、語の複数の声調、及び平叙文と疑問文との対比を表すような実際のイントネーション（語の声調を除く）を明らかにするものである。伝統的に第一声、第二声、第三声、第四声と呼ばれ（声調１から４）、その各々が他と区別される独自の特徴を持った４つの語の声調と、このような顕著な特徴のない中立声調（声調０）とがある。
【０００３】
声調の種類は中国語の音節を直接に構成する要素である。例えば、「ｍａ」は声調の種類によって以下の５つの異なる意味を持つ。
【０００４】
【数１】
ｍａ１（妈　：母）
ｍａ２（麻　：麻痺した）
ｍａ３（马　：馬）
ｍａ４（骂　：呪う）
ｍａ０（吗　：疑問の不変化詞）

このために重要な問題が生じる。テキストートゥースピーチ（ｔｅｘｔ－ｔｏ－ｓｐｅｅｃｈ：ＴＴＳ）合成においてイントネーションを合成する際に、語の声調と実際のイントネーションとの相互作用をどのように明らかにするか、ということである。これはＴＴＳを会話システムに適用する際に非常に重要である。会話システムでは例えば、疑問、メッセージの確認、及び感情が、人間によって、通常は音節のイントネーション（すなわち語の声調）と区別され、さらに通常の平叙文とも区別されるイントネーションのパターンで実現される［非特許文献１参照］。
【０００５】
これに対してとり得る解決策はおそらく、Ｆ０輪郭をアクセントと句の成分とに分解するフジサキのモデルであろう［非特許文献２参照］。イントネーションの変化をアクセントと句の成分との両者に分配してもよいが、モデルのパラメータ数は限られている。実際のイントネーションが語の声調に及ぼす影響に対処するため、言語学者は一般に音節［非特許文献３］または句［非特許文献４］のレベルでのピッチ範囲の変化に注目する。
【非特許文献１】Ｇ．コチャンスキー及びＣ．シー、「ソフトテンプレートを用いた韻律学モデリング」音声コミュニケーション、第３９巻、ｐｐ．３１１－３５２、２００３年（G. Kochanski and C. Shih, "Prosody modeling with soft templates," Speech Communication, Vol. 39, pp. 311-352, 2003.）
【非特許文献２】H.フジサキ及びK.ヒロセ、「日本語宣言文における音声基本周波数輪郭の分析」日本音響学会誌、第５巻、第４号、ｐｐ．２３３－２４２、１９８４年（H. Fujisaki and K. Hirose, "Analysis of voice fundamental frequency contours for declarative sentences of Japanese," J. Acoust. Soc. Japan, Vol.

5, No.4, pp. 233-242, 1984.）
【非特許文献３】Ｊ．シェン、「北京方言における声調とイントネーションのピッチ範囲」、実験的音声学における調査報告書、Ｔ．リン及びＬ．Ｊ．ワン編、北京大学出版局、ｐｐ．７３－１３０、１９８５年（中国語）（J. Shen, "Pitch range of tone and intonation in Beijing dialect," in Working papers in experimental phonetics, ed. by T. Lin and L. J. Wang, Beijing Univ. Press, pp. 73-130, 1985. (in Chinese)）
【非特許文献４】Ｚ．ウー、「標準中国語のためのイントネーション分析の新方法：文中の句輪郭の周波数転位処理」話し言葉の分析、知覚及び処理、Ｇ．ファンら編、ｐｐ．２５５－２６８、１９９６年（Z. Wu, "A new method of intonation analysis for standard Chinese: frequency transposition processing of phrasal contours in a sentence," Analysis, perception and processing of spoken language, ed. by G. Fant, et al, pp. 255-268, 1996.）
【非特許文献５】Ｙ．Ｒ．チャオ、中国語話し言葉の文法。バークレー、カリフォルニア大学出版局、１９６８年（Y. R. Chao, A grammar of spoken Chinese. Berkeley, University of California Press, 1968.）
【非特許文献６】Ｐ．クラトチヴィル、北京語のイントネーション、イントネーションシステム、２０ヶ国語の調査内、Ｄ．ハースト及びＡ．Ｄ．クリスト編、ケンブリッジ大学出版局、４１７－４３１、１９９８年（P. Kratochvil, Intonation in Beijing Chinese, in Intonation systems, a survey of twenty languages, ed. by D. Hirst and A. D. Cristo, Cambridge Uni. Press, 417-431, 1998.）
【非特許文献７】Ｊ．ニ及びＫ．ヒロセ、「標準中国語文の基本周波数輪郭の機能的モデリングの実験的評価」ＩＳＣＳＬＰ２０００、北京、ｐｐ．３１９－３２２、２０００年（J. Ni and K. Hirose, "Experimental evaluation of a functional modeling of fundamental frequency contours of standard Chinese sentences," ISCSLP2000, Beijing, pp. 319-322, 2000.）
【非特許文献８】Ｊ．ニ及びＨ．カワイ、「ピッチ範囲が中国語の声調とイントネーションパターンを固定する」音声韻律学２００４、奈良、ｐｐ．９５－９８、２００４年（J. Ni and H. Kawai, "Pitch targets anchor Chinese tone and intonation patterns," Speech Prosody 2004, Nara, pp. 95-98, 2004.）
【非特許文献９】Ｊ．ニ及びＨ．カワイ、「パラメトリックモデリング及び合成による分析ベースのパターンマッチングを通した声調特徴量の抽出」ＩＣＡＳＳＰ２００３、ｐｐ．７２－７５、２００３年（J. Ni and H. Kawai, "Tone feature extraction through parametric modeling and analysis-by-synthesis-based pattern matching," ICASSP2003, pp. 72-75, 2003）
【非特許文献１０】Ｊ．ニ及びＨ．カワイ、「関数モデル及びその評価による中国語基本周波数輪郭の骨格化」ＴＡＬ２００４、ｐｐ．１５１－１５４、北京、２００４年（J. Ni and H. Kawai, "Skeletonising Chinese fundamental frequency contours with a functional model and its evaluation," TAL2004, pp. 151-154, Beijing, 2004.）
【非特許文献１１】Ｊ．トゥハート、Ｒ．コリナー及びＣ．コーエン、イントネーションの知覚的研究：音声のメロディに対する実験的、音声学的アプローチ、ケンブリッジ大学出版局、１９９０年（J. 'tHart, R. Collier and A. Cohen, A perceptual study of intonation: an experimental-phonetic approach to speech melody, Cambridge University Press, 1990.）
【発明の開示】
【発明が解決しようとする課題】
【０００６】
このようなアプローチの限界は、測定されたピッチ範囲が多少とも語の声調の影響を含んでいることである。さらに、もしある発話中の語の声調がたまたま全て声調１であった

場合、ピッチ範囲の計算ができなくなる。というのも、声調１は高音域レベルの特性を有し、ピッチ範囲を推定するのに基準として利用可能な低音域の特徴がないからである。

【０００７】

この発明は、このイントネーションの変化を測定するという問題に別の方向から取組み、分離された個々の音節からの参考値の内部での声調変化を含む、声調の種類への依存性と、Ｆ０輪郭の起伏とを分解する際に生じる困難さを避けるようにする。

【０００８】

従って、この発明の目的の１つは、自然な条件下で、音声の基にあるイントネーションの変化を測定可能な方法を提供することである。

【０００９】

この発明の別の目的は、語の声調に影響されることなく、音声の基にあるイントネーションの変化を測定可能な方法を提供することである。

【課題を解決するための手段】

【００１０】

この発明の第１の局面に従えば、イントネーションの種類を声調の変形により特徴づける方法は、話者の個々の音節から得た語の声調の各々について、基本周波数（Ｆ０）ターゲットに関する参考値の所定の組を準備するステップを含み、Ｆ０ターゲットの参考値の組は対応する語の声調を特徴づけるものであり、話者のサンプル音声データ中の各音節についてＦ０ターゲット値を抽出するステップと、サンプル音声データ中の各音節のＦ０ターゲット値の各々について、その音節の語の声調に関する参考値から当該Ｆ０ターゲット値への変化の度合いを表す所定の第１のパラメータを計算するステップとをさらに含み、前記準備するステップは、語の声調の各々について前記話者による複数個の個々の音節を録音するステップと、それぞれの語の声調に従って、録音された個々の音節のＦ０ターゲット値を抽出するステップと、語の声調の各々について、語の声調を特徴づけるＦ０ターゲットの各々のＦ０ターゲット値を平均して前記参考値を求めるステップとを含む。

【００１２】

より好ましくは、この方法は、所定の第２のパラメータの分布が所定の第２のパラメータの所定の基準値の両側でつりあうように、前記所定の第１のパラメータを所定の第２のパラメータに正規化するステップをさらに含む。

【００１３】

この発明の第２の局面は、コンピュータ上で実行されると、上記したいずれかの全てのステップをコンピュータに行わせる、コンピュータプログラムに関する。

【発明を実施するための最良の形態】

【００１４】

Ａ．方法の概観

Ａ．１　変形

非特許文献７で扱われている、機能モデルで構築された変形は、さまざまな声域でのＦ０輪郭をλ時空間と呼ばれる正規化された空間にマッピングすることを可能にする。ここで、ｆ０はヘルツ表示のＦ０を表すものとし、λはλ（正規化された周波数）でのＦ０を表すものとする。ｆ０とλとの間の変形は以下の式で表される。

【００１５】

【数２】

$$\frac{\ln f0 - \ln f0_b}{\ln f0_t - \ln f0_b} = \frac{A(\lambda,\zeta) - A(\lambda_b,\zeta)}{A(\lambda_t,\zeta) - A(\lambda_b,\zeta)}, \qquad (1)$$

ここでＡ（λ，ζ）は単純な共振システム内での振幅－周波数応答を表す。

【００１６】

【数３】

$$A(\lambda,\zeta)=\frac{1}{\sqrt{\left(1-\left(1-2\zeta^2\right)\lambda\right)^2+4\zeta^2\left(1-2\zeta^2\right)\lambda}},\lambda\geq 1. \qquad (2)$$

ζは共振システムの減衰比を表す。物理的には、減衰比は共振システム中の粘性抵抗の等価物を表す。他のモデルパラメータは以下を示す。

【００１７】

[ｆ０$_b$，ｆ０$_t$]：声域の最高周波数と最低周波数

[λ$_b$，λ$_t$]：λで表した声域の最高周波数と最低周波数

声域[ｆ０$_b$，ｆ０$_t$]は話者に依存する。実際には、対象となる話者の発話の周波数範囲として測定することができる。ほとんどの場合、λ$_t$とλ$_b$とはそれぞれ１及び２に固定できる。

【００１８】

λとζとが与えられると、ｆ０は上述の変換で直接計算できる。便宜上、Ｔ$_{f0}$（）はζにおけるλからｆ０への変形を示すものとする。

【００１９】

ｆ０＝Ｔ$_{f0}$（λ，ζ）　　（３）

他方で、λ（又はζ）は、ｆ０とζ（又はλ）が与えられれば、反復処理によって決定することもできる。Ｔ$_λ$（）がζでのｆ０からλへの変形を表すものとする。ｆ０が大きくなるほど、λで表した値は小さくなる。

【００２０】

λ＝Ｔ$_λ$（ｆ０，ζ）　　（４）

さらに、Ｔ$_ζ$（）がλからｆ０への変形のためのζを表すものとする。

【００２１】

ζ＝Ｔ$_ζ$（λ，ｆ０）　　（５）

Ａ．２　声調の変形

この変換により、以下のζで示すように、[ｆ０$_b$，ｆ０$_t$]内でのｆ０$_1$からｆ０$_2$への変化を測定する方法が提供される。

【００２２】

ζ＝Ｔ$_ζ$（Ｔ$_λ$（ｆ０$_1$，ζ$_0$），ｆ０$_2$）　　（６）

ここでζ$_0$は、ｆ０$_1$及びｆ０$_2$をともにλ値にマッピングするときのζの基準値である。好ましくは、ζ$_0$は０．１５６に固定される。

【００２３】

ｆ０$_1$及びｆ０$_2$間の一対一のマッピングを保証するために、ζは（０，０．７]の集合に属していなければならない。これにより、以下のｆ０$_1$＝Ｔ$_{f0}$（λ$_i$，ζ$_0$）という条件下で図１に見られるように、個々のζについて、ｆ０$_1$及びｆ０$_2$間での制約が導かれる。

【００２４】

λ$_2$＝Ｔ$_λ$（Ｔ$_{f0}$（λ$_1$，ζ$_0$），ζ）　　（７）

ζが基準のζ$_0$（＝０．１５６）から遠ざかるにつれて、λ$_1$は非線形にかつ単調にλ$_2$へと変化し、その範囲は領域［１，２］の両端において急激に狭くなる。

【００２５】

ζをζ$_0$の両側でつりあわせるため、正規化された減衰比ζ$_n$をζ$_n$∈［－１，１］として次のように定義する。

【００２６】

【数４】

$$\zeta_n=\begin{cases}(\zeta-\zeta_0)/(0.7-\zeta_0), & for\ \zeta\geq\zeta_0,\\ (\zeta-\zeta_0)/\zeta_0, & for\ \zeta<\zeta_0.\end{cases} \qquad (8)$$

この方法を拡張して、語の声調及びピッチアクセント等の、２個のＦ０ターゲットのシ

ーケンス間の変化を測定することが可能である。ある声調の中でのすべてのＦ０ターゲットは、同じζ_0におけるλによる相対量として表される。この方法を２個の声調間の変化を測定するために用いる利点は、声調内の内部変化が見え、このため、実際の声調の変化を測定可能となることである。

【００２７】

図２から図４はこの声調変形をマンダリン語の声調に適用した例を示す。図２（ａ）は４個の語の声調（ボックス３０に示すように、声調１から声調４を同じ時間軸上で重ねたもの）を６回繰返した様子を示し、図２（ｂ）は$\zeta_n=0$を示し、これはターゲット声調変化がない、基準となる語の声調を表す。図３（ｂ）に示すように、ζ_nが２秒間に０から－1まで線形に変化すると、図２（ａ）の声調のシーケンスは図３（ａ）に示すものへと変化する。ζ_nは図４（ｂ）の太線に対応し、図２（ａ）の声調シーケンスは図４（ａ）に示す太線へと変化する。確かに、声域の非常に高い／低い領域ではピッチ範囲が狭くなる現象が実際の発声でよく見られる。

【００２８】

Ａ．３　イントネーションの変化測定

音節のイントネーションは声調と呼ばれる。音節と一致する時間－Ｆ０輪郭は声調パターンとして知られている。チャオ（Ｃｈａｏ）の声調理論［非特許文献５を参照されたい。］に従って、４つの語の声調を４個の声調パターンとして表し、さらにこれを、図５に示すようないくつかの選択されたＦ０ターゲットにより表す。各声調は主要ターゲットによって特徴づけられる［非特許文献６を参照されたい。］。図５では主要ターゲットを黒丸で示す。

【００２９】

Ｆ０輪郭で明示される声調の変化は、基となる語の声調を特定の態様で変更したものである［非特許文献６を参照］。Ｆ０輪郭は、Ｆ０ターゲットのシーケンスで信頼性をもって表すことができ、Ｆ０ターゲットの数と種類とは、声調パターンに従い、基となる語の声調から決定できる［非特許文献８を参照］。従って、声調変形を用いてＦ０輪郭から声調の変化を測定するアルゴリズムは、基本的に以下のステップを含む。

【００３０】

・初期化：話者による個々の音節から測定された平均のＦ０ターゲットに従って、４つの声調パターンについてＦ０ターゲットの基準値（参考値）を決定する。

【００３１】

・ステップ１：図５の声調パターンに従って、Ｆ０輪郭からＦ０ターゲット（観測値）を抽出する。Ｆ０輪郭からＦ０ターゲットを推定するためのアルゴリズムを、非特許文献９及び１０に記載のとおり利用することができ、これによってまず声調特徴を抽出し、その後これをＦ０ターゲットに変換する。

【００３２】

・ステップ２：声調パターンについて対（$f0_i$，$\hat{}f0_i$）を作成する。ここで、ｆ０ｉはｉ番目のＦ０ターゲットの観測値を表し、$\hat{}f0_i$（「ｆ」の前の「＾」記号は本来ｆの上部に表記すべきものである。）はその参考値を表す。声調０については、このＦ０ターゲットの参考値は、単に先行する声調での最後のＦ０ターゲットの参考値をとるものとする。

【００３３】

・ステップ３：$\zeta_i=T_\zeta(T_\lambda(\hat{}f0_i,\zeta_0),f0_i)$、及び$\zeta_n$を計算する。ただし，ｉ＝１，…Ｎ（Ｆ０ターゲットの数）とする。これがイントネーションの変化の特徴を表している。

【００３４】

図６は、（ａ）ζ_n（丸）により特徴が表されたイントネーションパターンの推定に用いられたＦ０ターゲット対と、（ｂ）対応する発話データで得られたＦ０輪郭のためのＦ０ターゲット対との、参考値（三角）と観測値（丸）とをプロットしている。線Ｐ０Ｐ４は$\zeta_n=-1.045t+0.686$を示し、線Ｐ５Ｐ７は$\zeta_n=-0.809t+1.$

１９８を示す。
【００３５】
Ｂ．実施例の説明
Ｂ．１　構造
Ｂ．１．１　機能ブロック
図７はこの発明の一実施例に従った音声合成システム４０を示すブロック図である。図７を参照して、音声合成システム４０は、所定の話者の基準発話のための記憶装置５０と、話者のサンプル発話を記憶するための記憶装置５２と、基準発話の声調の各々に対する基準Ｆ０ターゲットを抽出し、さらに記憶装置５２に記憶されたサンプル発話の各々について、イントネーション変化を示す正規化された減衰比ζ_nのシーケンスを抽出するためのイントネーション抽出モジュール５４とを含む。
【００３６】
音声合成システム４０はさらに、基準発話の基準Ｆ０ターゲットを記憶するための記憶装置５６と、ζ_nのシーケンスを記憶するための記憶装置５８とを含む。減衰比ζ_nのシーケンスは、サンプル発話のイントネーション変化の特徴を表すものである。従って、ユーザは、記憶装置５８に記憶されたζ_nのシーケンスを利用して、所望のイントネーションを指定することができる。
【００３７】
音声合成システム４０はさらに、合成すべき入力テキスト６２と関連付けられたイントネーション情報６０を受け、入力テキスト６２中の音節の各々についてＦ０を合成するためのＦ０シンセサイザ６４と、入力されたテキスト６２とＦ０シンセサイザ６４から出力されたＦ０とに従って音声信号を合成するための音声シンセサイザ６６とを含む。
【００３８】
イントネーション抽出モジュール５４は、記憶装置５０内の基準発話の音節の各々からＦ０ターゲットを抽出し、抽出されたｆ０ターゲットを記憶装置５６に記憶するための第１のターゲット抽出モジュール８０と、記憶装置５２内のサンプル発話の音節の各々からＦ０ターゲットを抽出するための第２のターゲット抽出モジュール８２と、第２のターゲット抽出モジュール８２から出力されたＦ０ターゲットの各々について、減衰比ζ_nを計算し、ζ_nのシーケンスを記憶装置５８に出力するためのζ_n計算モジュール８４とを含む。
【００３９】
Ｆ０シンセサイザ６４は、イントネーション情報内のζ_nのシーケンスからζを計算するζ計算モジュール９０と、以下の式に従って、入力テキスト６２の各々の音節の$f0_i$を計算し、計算された$f0_i$を音声シンセサイザ６６に出力するためのＦ０計算モジュール９０とを含む。
【００４０】
$$f0_i = T_{f0}(T_\lambda(f0_i, \zeta_0), \zeta) \qquad (9)$$
Ｂ．１．２　コンピュータによる実現
図７に示されたモジュールは、この実施例ではコンピュータソフトウェアで実現される。図８は第１のターゲット抽出モジュール８０を実現するコンピュータプログラムの制御構造を示す。図８を参照して、プログラムはステップ１００で始まり、基準発話に見出される声調１～声調４の各々について、ステップ１０２～１２０が繰返される。
【００４１】
ステップ１０２で、変数ＳＵＭがゼロに初期化される。
【００４２】
ステップ１１０で、基準発話内の、関心のある声調データの全てについて、ステップ１１２～１１６が繰返される。ステップ１１４で、音節の音声データからＦ０ターゲットが抽出される。抽出されたＦ０はステップ１１６でＳＵＭに加えられる。
【００４３】
ステップ１１２から１１６が関心のある声調の音節全てに対し繰返された後、ステップ

１１８でＳＵＭの平均を求める。ステップ１２０で、この平均が、対象の声調と関連付けた上でメモリに記憶される。

【００４４】
この処理の終わりには、声調１～声調４の平均Ｆ０がメモリに記憶されていることになる。

【００４５】
図９は図７に示す第２のターゲット抽出モジュール８２及びζ_n計算モジュール８４を実現するコンピュータプログラムの制御構造を示す。図９を参照して、ステップ１４０で、記憶装置５２に記憶されたサンプル発話の全てについてＦ０輪郭が計算される。ステップ１４２で、入力テキスト６２（図７を参照）の全ての音節について、ステップ１４４から１５２が繰返される。

【００４６】
この繰返しでは、まず、処理中の音節の声調のＦ０ターゲットが抽出される。抽出されたｉ番目のＦ０ターゲットを$f0_i$，$1 \leqq i \leqq N$（発話中のターゲットの数）とする。

【００４７】
ステップ１４６で、ステップ１４４で抽出された$f0_i$が音節の声調パターンの$\wedge f0_i$と対にされる。ここで$\wedge f0_i$は$f0_i$の参考値を表す。声調０については、そのＦ０ターゲットの参考値は単に、先行する声調の最後のＦ０ターゲットの参考値をとるだけである。

【００４８】
ステップ１４８で、ζ_iが以下の式に従って計算される。

【００４９】
$$\zeta_i = T_\zeta(T_\lambda(\wedge f0_i, \zeta_0), f0_i) \qquad (10)$$
ステップ１５０で、正規化されたζ_{ni}（$1 \leqq i \leqq N$）が以下の式に従って計算される。

【００５０】
【数５】
$$\zeta_{ni} = \begin{cases} (\zeta_i - \zeta_0)/(0.7 - \zeta_0), & for \zeta_i \geq \zeta_0 \\ (\zeta_i - \zeta_0)/\zeta_0 & for \zeta_i < \zeta_0 \end{cases} \qquad (11)$$
ステップ１５２で、結果ζ_{ni}が記憶装置５８に記憶される（図７を参照）。

【００５１】
記憶装置５２に記憶されているサンプル発話の音節全てについて上述の処理を繰返した後、ユーザは正規化されたζ_nを用いればどのようなイントネーションも記述できる。従って、イントネーション情報６０はζ_nのシーケンスの形で準備することができる。

【００５２】
この実施例では、図７に示すＦ０シンセサイザ６４もまたコンピュータソフトウェアで実現される。このコンピュータプログラムの制御構造を図１０に示す。

【００５３】
図１０を参照して、Ｆ０シンセサイザ６４が起動されると、まずイントネーション情報６０内のイントネーションデータζ_{ni}を読出す。次に、ステップ１７２で、入力テキスト６２の音節全てについてステップ１７４から１７８を繰返す。ここでζ_{ni}（$1 \leqq i \leqq N$）はイントネーション情報６０の正規化された減衰率のシーケンスとする。

【００５４】
ステップ１７４で、式（１１）の逆関数に従って、ζ_{ni}からζ_iを計算する。

【００５５】
ステップ１７６で、ｉ番目の音節（声調）のＦ０ターゲット$f0_i$が以下の式に従って計算される。

【００５６】
$f0_i = T_{f0}(T_\lambda(\hat{}f0_i, \zeta_0), \zeta_i)$　　（１２）
ここで$\hat{}f0_i$は基準発話から抽出された参考値（Ｆ０ターゲット）を表し、ζ_0は定数（好ましくは、ζ_0は０．１５６）を表す。
【００５７】
ステップ１７８で、このようにして計算された$f0_i$がメモリに記憶される。
【００５８】
入力テキスト６２の全ての音節について、ステップ１７４から１７８が繰返された後、イントネーション情報６０によりイントネーションパターンが指定された入力テキスト６２中の声調のシーケンスのＦ０ターゲットとして、$f0_i$のシーケンスがステップ１８０で出力される。
【００５９】
Ｂ．１．３　コンピュータハードウェア
図１１は上述のコンピュータプログラムを実行するこの実施例のコンピュータシステム３３０の外観を示し、図１２はこのシステム３３０をブロック図で示す。
【００６０】
図１１を参照して、このコンピュータシステム３３０は、ＦＤ（フレキシブルディスク）ドライブ３５２およびＣＤ－ＲＯＭ（コンパクトディスク読出専用メモリ）ドライブ３５０を有するコンピュータ３４０と、キーボード３４６と、マウス３４８と、モニタ３４２と、一対のスピーカ３７２と、マイクロフォン３７０と、を含む。
【００６１】
図１２を参照して、コンピュータ３４０はさらに、ＣＰＵ（中央処理装置）３５６と、ＣＰＵ３５６、ＦＤドライブ３５２およびＣＤ－ＲＯＭドライブ３５０に接続されたバス３６６と、ハードディスク３５４と、ブートアッププログラム等を記憶する読出専用メモリ（ＲＯＭ）３５８と、ＣＰＵ３５６に接続され、アプリケーションプログラム命令、システムプログラム、及びデータ等を記憶するランダムアクセスメモリ（ＲＡＭ）３６０とを含む。
【００６２】
ここでは示さないが、コンピュータ３４０はさらにローカルエリアネットワーク（ＬＡＮ）への接続を提供するネットワークアダプタボードを含んでもよい。
【００６３】
コンピュータシステム３３０に上述の音声合成システムを実現させるためのコンピュータプログラムは、ＣＤ－ＲＯＭドライブ３５０またはＦＤドライブ３５２に挿入されるＣＤ－ＲＯＭ３６２またはＦＤ３６４に記憶され、さらにハードディスク３５４に転送される。または、プログラムは図示しないネットワークを通じてコンピュータ３４０に送信されハードディスク３５４に記憶されてもよい。プログラムは実行の際にＲＡＭ３６０にロードされる。ＣＤ－ＲＯＭ３６２から、ＦＤ３６４から、またはネットワークを介して、直接にＲＡＭ３６０にプログラムをロードしてもよい。
【００６４】
図８から図１０を参照して説明したこのプログラムは、コンピュータ３４０にこの実施例の音声合成システム４０の機能ブロックを実現させるための複数の命令を含む。この方法を行なわせるのに必要な基本的機能のいくつかはコンピュータ３４０上で動作するオペレーティングシステム（ＯＳ）またはコンピュータ３４０にインストールされるサードパーティのプログラムにより提供される。従って、このプログラムはこの実施の形態のシステムおよび方法を実現するのに必要な機能全てを必ずしも含まなくてよい。このプログラムは、命令のうち、所望の結果が得られるように制御されたやり方で適切な関数または「ツール」を呼出すことにより、上述の処理を行う命令のみを含んでいてもよい。コンピュータシステム３３０の動作は周知であるので、ここでは繰返さない。
【００６５】
Ｂ．２　動作

この実施例の、上述の音声合成システム４０（図７を参照）は以下のように動作する。音声合成システム４０の動作は３段階である。すなわち、基準発話からのＦ０ターゲットの抽出と、基準発話からのζ_nの計算と、Ｆ０ターゲット及び音声合成とである。これらの段階における音声合成システム４０の動作を以下で説明する。

【００６６】

Ｂ．２．１　基準発話からのＦ０ターゲットの抽出

図７を参照して、所定の話者の音声データを、声調１～声調４の全てについて録音し、基準発話として記憶装置５０に記憶する。声調１～声調４の各々について、第１のターゲット抽出モジュール８０により、基準発話からＦ０ターゲットが抽出される。声調１～声調４の各々について平均のＦ０ターゲットが記憶装置５６に記憶される。

【００６７】

Ｂ．２．２　基準発話からのζ_nの計算

基準発話と同じ話者のサンプル発話を録音し、記憶装置５２に記憶する。サンプル発話の各々の各音節について、第２のターゲット抽出モジュール８２がＦ０ターゲットを抽出する。その後、モジュール８２から出力されたＦ０ターゲットの各々について、ζ_n計算モジュール８４がζ_nを計算し、サンプル発話の各々についてζ_nのシーケンスを生成する。

【００６８】

Ｂ．２．３　Ｆ０ターゲット及び音声合成

ユーザは、入力テキスト６２と、入力テキストをそのイントネーションで合成したいと考えているイントネーションを特定する関連のイントネーション情報６０とを準備する。ユーザは、記憶装置５８に記憶されているζ_nのシーケンスを調べることにより、イントネーション情報を準備することができる。

【００６９】

イントネーション情報６０と入力テキスト６２とが準備されると、入力テキスト６２の各音節について、ζ計算モジュール９０がζを計算し、これをＦ０計算モジュール９２に出力する。例えば、ｉ番目の音節に対し、ζ計算モジュール９０は式（１１）の逆関数に従って$\zeta_{n\,i}$からこの音節のζ_iを計算する。

【００７０】

Ｆ０計算モジュール９２は、音節の各々に対し、このようにして計算されたζ_iと、記憶装置５６に記憶された$\hat{}f0_i$と、定数ζ_0＝０．１５６とに以下の関数を適用してＦ０ターゲット$f0_i$を計算する。

【００７１】

$$f0_i = T_{f0}(T_{\lambda}(\hat{}f0_i, \zeta_0), \zeta_i) \qquad (13)$$

この結果、入力テキスト６２内の音節について、Ｆ０計算モジュール９２により、$f0_i$のシーケンスが出力される。このシーケンスが音声シンセサイザ６６に与えられる。

【００７２】

Ｆ０計算モジュール９２から$f0_i$のシーケンスが与えられると、音声シンセサイザ６６は、イントネーション情報６０で指定されたイントネーションを備えた入力テキスト６２の音声信号６８を合成することができる。

【００７３】

Ｃ．実験結果

ここで提案した方法が、測定されたＦ０輪郭内の、語の声調よりも高いレベルのイントネーションの変化を明らかにすることが可能であると示すために、２つの実験結果を報告する。音声サンプルは中国語音声コーパスから選択され、専門のナレータに朗読してもらった。ナレータの声域[$f0_b$，$f0_t$]は[１００Ｈｚ，５００Ｈｚ]と一致し、ナレータによる語の声調の参考値は表１に示されるとおりである。太字は主要ターゲットを示す。これらの参考値に対応する声調パターンを図２（ａ）に見ることができる。

【００７４】

【表１】

表1. ナレータの語の声調の参考値

	不完全指定	参考値 (Hz)
声調1	<**P11**, P12>	<**370**, 363>
声調2	<V21, **P22**>	<250, **355**>
声調3	<P31, **V31**, P32>	<235, **200**, 280>
声調4	<**P41**, V41>	<**375**, 225>

【００７５】

【表２】

表2. 図13(a)のイントネーション変化のパラメータ

i	音節	$f0_i$	$\hat{f}0_i$	ζ_i	ζ_{ni}
1	ni3	243	250 (V21)	0.149	−0.045
2	(声調2)*	367	**355 (P22)**	**0.169**	**0.024**
3	hao3	197	235 (P31)	0.115	−0.263
4	(声調3)*	152	**200 (V31)**	**0.090**	**−0.423**
5		204	280 (P32)	0.091	−0.417

*声調3の後に声調3が続くと、前者は声調2に変化する

図１３～図１６に示される結果は、４つの慣用の挨拶を含むイントネーション変化の分析から得られた。４つの挨拶の実際のイントネーションは音韻論的には同じであるが、語の声調のためにＦ０輪郭は大きく起伏する。計算の例として、表２は、図１３（ａ）に示されたサンプルからの観測値ｆ０$_i$，ｉ＝１，…５、対応の参考値＾ｆ０$_i$、及び結果として得られるパラメータζ$_i$及びζ$_{ni}$を列挙している。これらの結果は図１３（ｂ）に示される。

【００７６】

この例では、文のアクセントは、声調２の主要ターゲット（最初の声調３の表面声調）である０．０２４から第２の声調３の－０．４２３までζ$_n$が下降したことで示される。他の文の文アクセントもまた、基となる声調の種類に関わりなく一貫して下降するように思われる。この４つの挨拶で示される基本的な特徴は、（１）文のアクセントは発話の最後に位置し、もう１つの音節にかかること、（２）最後の声調（声調１～４）はその参考声調パターンを維持する（すなわちζ$_n$が変化しない）ことである。声調０は最後の非－声調０である声調の連続したものであるとみなされる。この結果は上述の仮定と一致する。イントネーション変化の現象は、例えば非特許文献１１で例示されているように、非声調言語でイントネーションを説明するのに通常用いられるいわゆる「ハットパターン」に非常に類似している。

【００７７】

図１７は声調及びイントネーションを合成する例を示す。図１７（ａ）は基となる語の声調の参考値を示す。図１７（ｂ）はζ$_n$（ｔ）によりイントネーションパターンをプロットする。図１７（ｃ）はこれらのＦ０ターゲット（丸）とこれらのターゲットによりモデルによって与えられる輪郭（連続線）とを示す。「＋」のシーケンスはサンプル発話の測定されたＦ０輪郭を示す。

【００７８】

図１７から明らかなように、モデルによって与えられるＦ０輪郭は元のＦ０輪郭に非常に近い。

【００７９】

図１８は同じ話者にいくつかの数字列を読んでもらうことで得られたさらなる結果を示す。朗読した数字列は、言語学的意味がないため、中立である。明瞭な結果を求めるため、主要な声調ターゲットのζ$_n$値のみを図にプロットする。加えて、これらの発話では休

止（ポーズ）がない。イントネーション変化には２つの形状が現れる。１つは最初から最後まで下がる線である（左側）。他方は、下降部とそれに続く平坦部とからなる線である。この下降は最初の２個の音節間で起こる。明らかになったイントネーション変化は、語の声調を越えた高いレベルで体系的である。

【００８０】

３人の話者による約２００個の中国語サンプルを分析した。これらのサンプルでは実際のイントネーションは多少変化するものの、分析した結果は、この方法により、上で示したとおりイントネーションの変化をはっきりと明らかにできることを示した。

【００８１】

Ｄ．結論

この発明の実施の形態は、測定されたＦ０輪郭から語の声調を除外したイントネーション変化を測定する方法に関する。イントネーション変化は語の声調パターンを構成する選択されたＦ０ターゲットを用いてサンプリングされ、時間軸上の１点のパラメータで特徴づけられる。実験結果から、この提案した方法が、Ｆ０輪郭に埋もれ、語の声調と混じりあった、実際のマンダリン語のイントネーションを分析するのに非常に有望であることがわかった。明らかにされた実際のイントネーションは、非声調言語で報告されたイントネーションとの類似性を示した。提案された方法は基となる語の声調をともなったＦ０輪郭の自動的な分析を試みるものであり、これは音声合成、認識、さらには理解において決定的に重要である。

【００８２】

今回開示された実施の形態は単に例示であって、本発明が上記した実施の形態のみに制限されるわけではない。本発明の範囲は、発明の詳細な説明の記載を参酌した上で、特許請求の範囲の各請求項によって示され、そこに記載された文言と均等の意味および範囲内でのすべての変更を含む。

【図面の簡単な説明】

【００８３】

【図１】λ_1、λ_2及びζ間の条件を示す図である。

【図２】声調変形をマンダリン語の声調に適用した例を示す図である。

【図３】声調変形をマンダリン語の声調に適用した別の例を示す図である。

【図４】声調変形をマンダリン語の声調に適用した別の例を示す図である。

【図５】Ｆ０ターゲットをベースとしてマンダリン語の声調を表す図である。

【図６】ζ_n（丸）でのイントネーション変化の推定に用いられるＦ０ターゲット対と、元のＦ０輪郭のための、参考値（三角）と観測値（丸）とをプロットした図である。

【図７】この発明の一実施の形態に従った音声合成システム４０のブロック図である。

【図８】第１のＦ０ターゲット抽出モジュール８０を実現するコンピュータプログラムの制御構造をフローチャートで示す図である。

【図９】第２のターゲット抽出モジュール８２とζ_n計算モジュール８４とを実現するコンピュータプログラムの制御構造をフローチャートで示す図である。

【図１０】Ｆ０シンセサイザ６４を実現するコンピュータプログラムの制御構造をフローチャートで示す図である。

【図１１】一実施の形態に係るコンピュータプログラムを実行するコンピュータシステム３３０の斜視図である。

【図１２】システム３３０のブロック図である。

【図１３】慣用の挨拶「ｎｉ３ｈａｏ３」（こんにちは）のＦ０輪郭を示す図である。

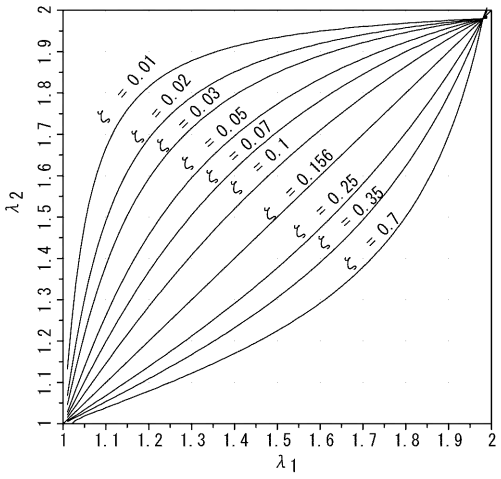
【図１４】慣用の挨拶「ｚｅｎ３ｍｅ０ｙａｎｇ４ａ０？」（いかがお過ごしですか）のＦ０輪郭を示す図である。

【図１５】慣用の挨拶「ｎｉ３ｍａｎｇ２ｍａ０？」（お忙しいですか）のＦ０輪郭を示す図である。

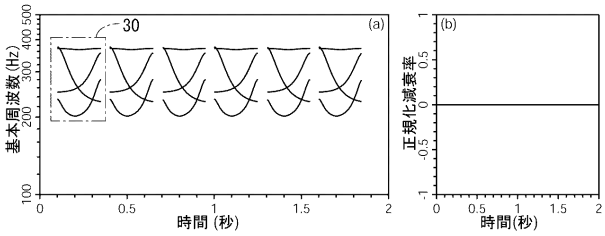
【図１６】慣用の挨拶「ｎｉ３ｓｈｅｎ１ｔｉ３ｈａｏ３ｍａ０？」（ごきげんいかがですか）のＦ０輪郭を示す図である。

【図１７】語による韻律の特徴と、語によらない韻律の特徴とを合成する例を示す図である。
【図１８】朗読された数字列での中立イントネーションの変化を示す図である。
【符号の説明】
【００８４】
４０　　音声合成システム
５０、５２、５６、５８　　記憶装置
５４　　イントネーション抽出モジュール
６０　　イントネーション情報
６２　　入力テキスト
６４　　Ｆ０シンセサイザ
６６　　音声シンセサイザ
６８　　イントネーションのある音声信号
８０　　第１のＦ０ターゲット抽出モジュール
８２　　第２のＦ０ターゲット抽出モジュール
８４　　ζ_n計算モジュール
９０　　ζ計算モジュール
９２　　Ｆ０計算モジュール

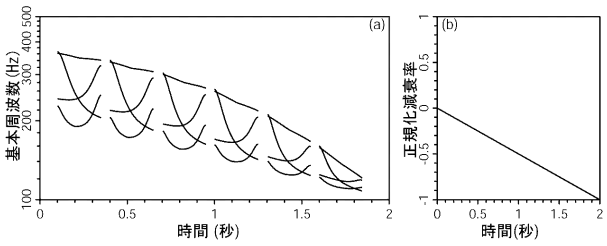
【図１】



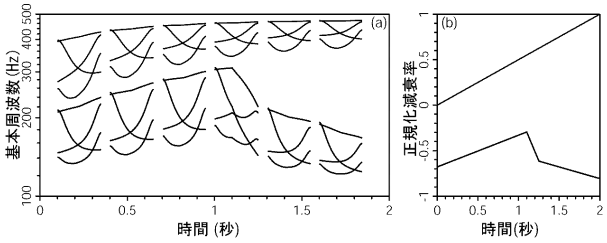
【図２】



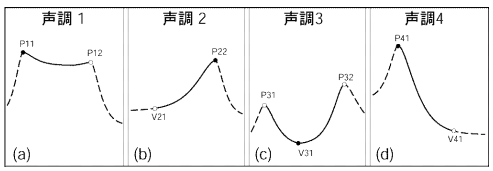
【図３】



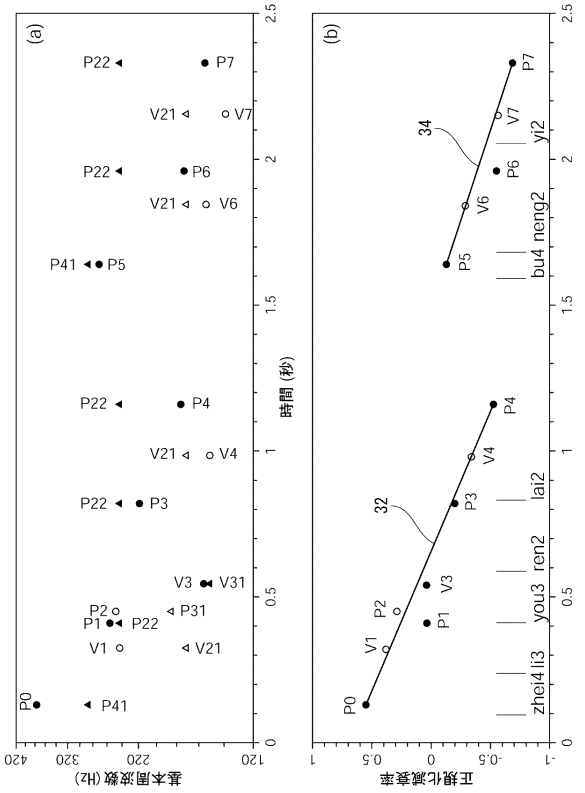
【図４】



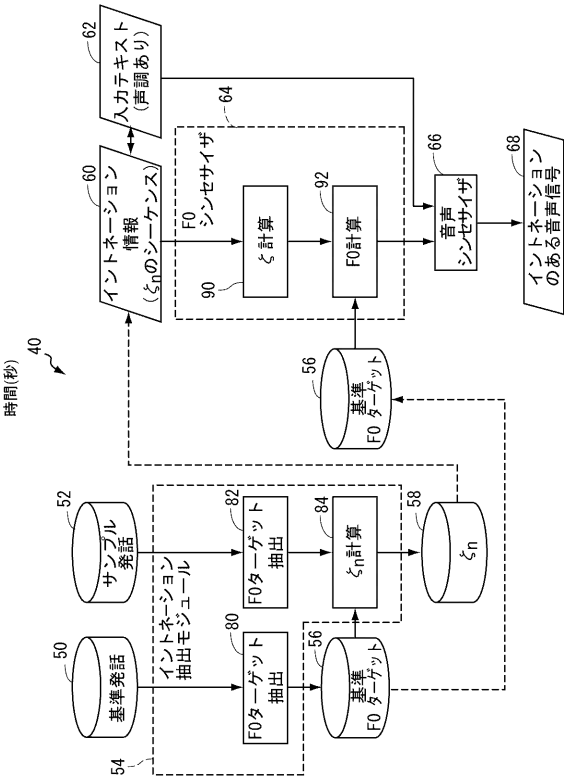
【図５】



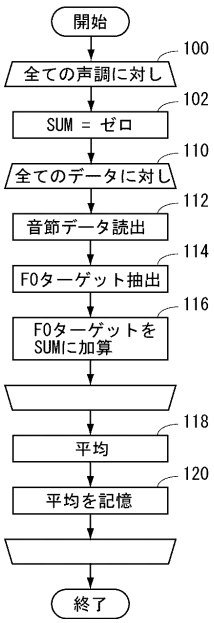
【図６】


(a)
(b)
P22
P7
V21
V7
P6
V6
P41
P5
P4
V4
P3
V3
V31
P2
P31
P1
V1
P0
時間（秒）
基本周波数（Hz）
正規化減衰率
yi2
neng2
lai2
ren2
you3
li3
zhei4
bu4

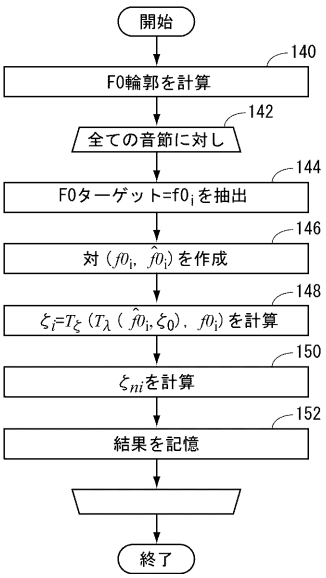
【図７】


62
入力テキスト（声調あり）
64
60
イントネーション情報（ζnのシーケンス）
F0シンセサイザ
ζ計算
90
92
F0計算
66
音声シンセサイザ
68
イントネーションのある音声信号
40
時間(秒)
56
基準F0ターゲット
52
サンプル発話
82
F0ターゲット抽出
84
ζn計算
58
ζn
イントネーション抽出モジュール
50
基準発話
80
F0ターゲット抽出
56
基準F0ターゲット
54

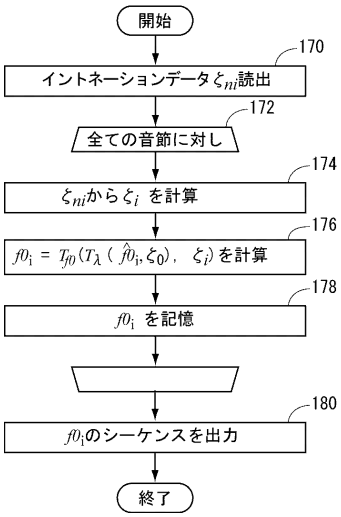
【図８】


開始
100
全ての声調に対し
102
SUM = ゼロ
110
全てのデータに対し
112
音節データ読出
114
F0ターゲット抽出
116
F0ターゲットをSUMに加算
118
平均
120
平均を記憶
終了

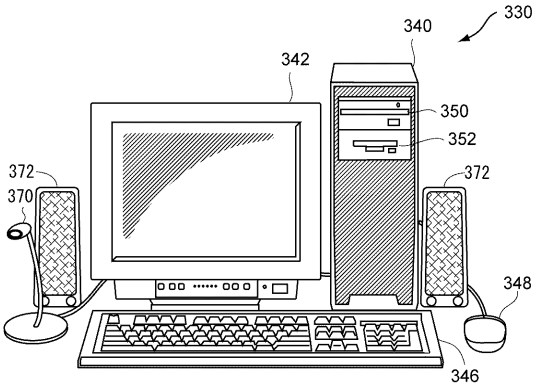
【図９】


開始
140
F0輪郭を計算
142
全ての音節に対し
144
F0ターゲット=f0i を抽出
146
対 (f0i, f̂0i) を作成
148
ζi=Tζ (Tλ (f̂0i,ζ0), f0i) を計算
150
ζniを計算
152
結果を記憶
終了

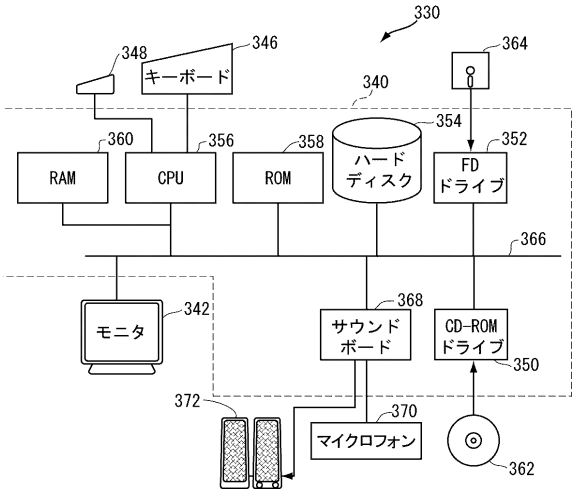
【図１０】



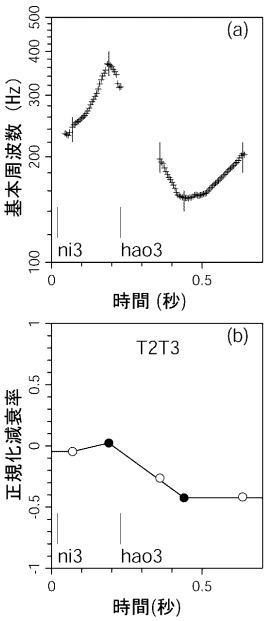
【図１１】



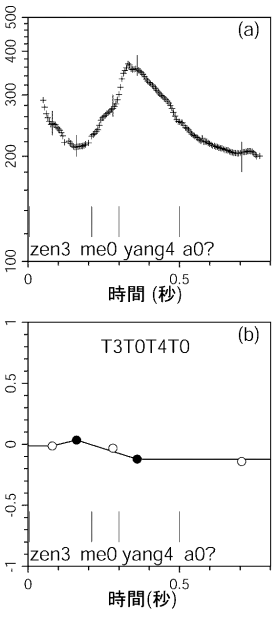
【図１２】



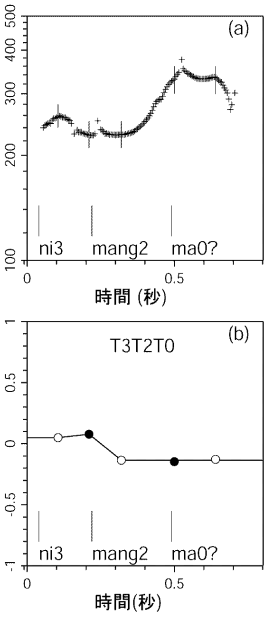
【図１３】



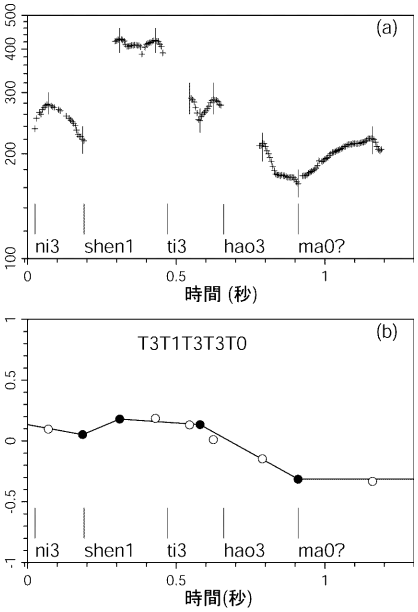
【図１４】



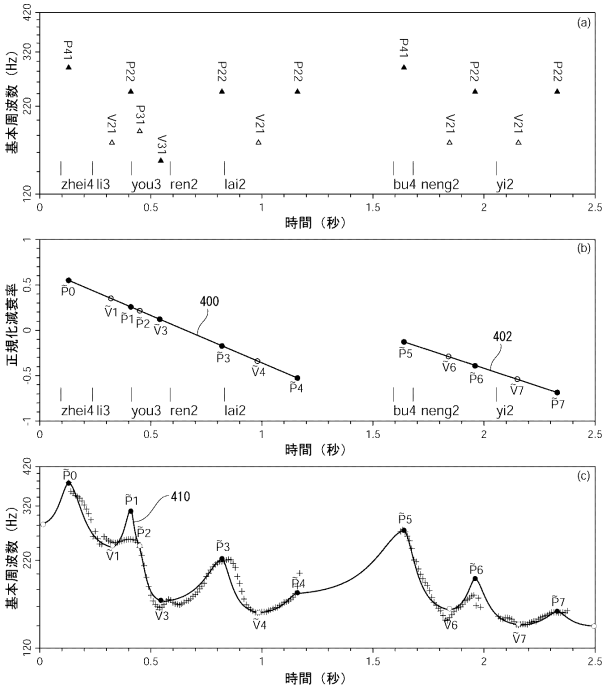
【図１５】



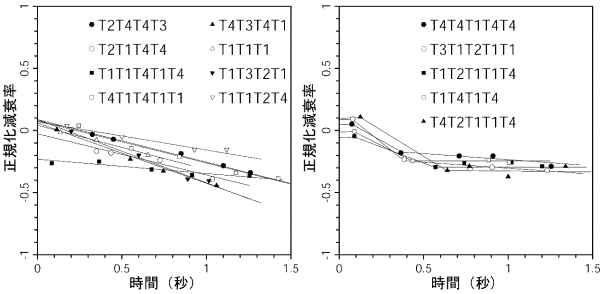
【図１６】



【図１７】



【図１８】



フロントページの続き

(56)参考文献 特開２００３－３３０４８２（ＪＰ，Ａ）
特開２００２－２２９５９０（ＪＰ，Ａ）
特開２００５－２６５９５５（ＪＰ，Ａ）
特開２００５－２５０２６４（ＪＰ，Ａ）
広瀬啓吉他，”基本周波数パターン生成過程のモデルによる標準中国語音声の韻律的特徴の分析と定式化”，日本音響学会誌，Vol.50，No.3（1994-03），pp.177-187
森大毅他，”単語レベルのＦ０レンジを考慮した中国語音声の韻律制御”，日本音響学会１９９９年秋季研究発表会講演論文集-Ｉ-，2-Q-20（1999-09），pp.319-320
徐大威他，”中国語単語のＦ０レンジの変化に対する許容度に関する検討”，日本音響学会２００１年秋季研究発表会講演論文集-Ｉ-，1-2-14（2001-10），pp.233-234
徐大威他，”中国語二音節単語の相対Ｆ０変化域の不変性”，電子情報通信学会技術研究報告，Vol.101，No.86，SP2001-13（2001-05），pp.29-34

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ１０Ｌ １１／００－１３／０８