

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第4617092号
(P4617092)

(45) 発行日 平成23年1月19日(2011.1.19)

(24) 登録日 平成22年10月29日(2010.10.29)

(51) Int.Cl. F I
G 1 0 L 15/02 (2006.01) G 1 0 L 15/02 3 0 0 A
G 1 0 L 13/08 (2006.01) G 1 0 L 13/08 1 2 7 C

請求項の数 9 外国語出願 (全 13 頁)

<p>(21) 出願番号 特願2004-74594 (P2004-74594) (22) 出願日 平成16年3月16日 (2004.3.16) (65) 公開番号 特開2005-265955 (P2005-265955A) (43) 公開日 平成17年9月29日 (2005.9.29) 審査請求日 平成19年2月26日 (2007.2.26)</p> <p>(出願人による申告) 平成15年9月30日付け「第年5E1 8号」平成15年度、支出負担行為担当官 総務省大臣官房会計課企画官、研究テーマ「携帯電話等を用いた多言語自動翻訳システム」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</p>	<p>(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 チョウ キンソン 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 中村 哲 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 広瀬 啓吉 神奈川県鎌倉市佐助1-7-15 審査官 毛利 太郎</p>
--	---

最終頁に続く

(54) 【発明の名称】 中国語の声調分類装置及び中国語のF0生成装置

(57) 【特許請求の範囲】

【請求項1】

中国語音声の声調核境界における基本周波数(F0)により声調を識別する特徴量を含む特徴量の組が与えられると、中国語で使用される声調分類の確率を出力するように、トレーニングデータセットを用いてトレーニングされた声調モデルを記憶するための手段を含む中国語の声調分類装置であって、前記声調核境界におけるF0により声調を識別する特徴量は先頭ギャップ、末尾ギャップ、及び、声調核のF0輪郭の傾斜勾配の符号の組み合わせを含み、

前記声調分類装置はさらに、

入力された中国語の音声データを一連の声調核にセグメント化するための手段と、

前記声調核の各々から声調核境界におけるF0により声調を識別する特徴量を抽出するための手段と、

前記抽出するための手段によって抽出された音響特徴量を前記声調モデルに適用し、前記声調モデルにより出力される最も高い確率を達成する声調分類を選択するためのパターンマッチング手段とを含む、中国語の声調分類装置。

【請求項2】

声調核境界におけるF0により声調を識別する前記特徴量はさらに、声調核の先頭F0及び末尾F0を含む、請求項1に記載の中国語の声調分類装置。

【請求項3】

声調核境界におけるF0により声調を識別する前記特徴量はさらに、声調核の正規化され

たパワーを含む、請求項 1 又は請求項 2 に記載の中国語の声調分類装置。

【請求項 4】

声調核境界における F 0 により声調を識別する特徴量を含む特徴量の組が与えられると、中国語で使用される声調分類の確率を出力するように、トレーニングデータセットを用いてトレーニングされた声調モデルを記憶するための手段と、

入力された中国語の音声データを一連の声調核にセグメント化するための手段と、

前記声調核の各々から声調核境界における F 0 により声調を識別する特徴量を抽出するための手段とを含む中国語の声調分類装置であって、声調核境界における F 0 により声調を識別する前記特徴量は、先頭ギャップ、末尾ギャップ、及び、前記声調核の F 0 輪郭の傾斜勾配の符号の組み合わせを含み、

10

前記声調分類装置はさらに、

前記抽出するための手段によって抽出された先頭ギャップ、末尾ギャップ、及び、前記声調核の F 0 輪郭の傾斜勾配の符号の組み合わせに従って声調の分類を判断し、判断された分類を出力するための声調分類手段を含む、中国語の声調分類装置。

【請求項 5】

前記声調分類手段は、声調の先頭ギャップが正であり末尾ギャップが正であるとき、その声調が中国語の第一声であると判断するための手段を含む、請求項 4 に記載の中国語の声調分類装置。

【請求項 6】

前記声調分類手段はさらに、声調の先頭ギャップが負であり末尾ギャップが正であるとき、その声調が中国語の第二声であると判断するための手段を含む、請求項 5 に記載の中国語の声調分類装置。

20

【請求項 7】

前記声調分類手段はさらに、声調の先頭ギャップが負であり末尾ギャップが負であるとき、その声調が中国語の第三声であると判断するための手段を含む、請求項 6 に記載の中国語の声調分類装置。

【請求項 8】

前記声調分類手段はさらに、声調の先頭ギャップが正であり末尾ギャップが負であるとき、その声調が中国語の第四声であると判断するための手段を含む、請求項 7 に記載の中国語の声調分類装置。

30

【請求項 9】

中国語の F 0 生成装置であって、

声調核境界における F 0 により声調を識別する特徴量を含む特徴量の組が与えられると、中国語で使用される声調分類の確率を出力するように、トレーニングデータセットを用いてトレーニングされた声調モデルを記憶するための手段を含む中国語の F 0 生成装置であって、関連付けにより声調を識別する前記特徴量は先頭ギャップ、末尾ギャップ、及び、声調核の F 0 輪郭の傾斜勾配の符号の組み合わせを含み、

前記 F 0 生成装置はさらに、

構文解析された中国語テキストが与えられると、その中国語テキスト内の音声単位の各々について可能な中国語声調の確率を出力する、確率的 F 0 モデルを記憶するための手段と、

40

前記確率的 F 0 モデルの出力に従って、入力された中国語のテキストに適合する F 0 のシーケンスを生成するための手段と、

前記生成するための手段によって出力された F 0 シーケンスが、前記声調モデルと整合しているか否かを判断するための手段とを含む、中国語の F 0 生成装置。

【発明の詳細な説明】

【技術分野】

【0001】

この発明は声調言語の声調分類に関し、特に、中国語等の声調言語において、文脈による声調の変化を非常に正確に識別することに関する。

50

【背景技術】

【0002】

中国語の基本的語彙声調（第一声、第二声、第三声、第四声と呼ばれる）は通常、それらの基本周波数（“F0”）輪郭のパターンが異なることによって特徴付けられる。すなわち、第一声は高いレベルの輪郭、第二声は途中から上がる輪郭、第三声は低く一旦下がる輪郭、第四声は高いところから下がる輪郭である。これらを識別するため、研究者はこれまでの研究において、第一声（高いF0）と第三声（低いF0）との区別（非特許文献1を参照されたい。）、及び第三声（低いF0）と第四声（高いF0）との区別（非特許文献2を参照されたい。）にはF0の高さが決定的に重要であると報告している。

【0003】

F0傾斜勾配は、第一声（平坦）、第二声（上がる）、及び第四声（下がる）等の、動きの方向が異なる声調間を区別するのに有効である（非特許文献2を参照されたい。）。 10

【0004】

声調のF0輪郭は、連続した音声においては、互いに分離された音節に比べてかなり変動する。F0の高さのみでなく、傾斜勾配も相当に変動するため、表面上のF0輪郭からその下にある声調性を特定することはできない。他方で、知覚実験によれば、人間は、声調的な文脈が与えられれば、F0がかなり変動しても、その下にあると思われる語彙的な声調を高い一貫性をもって知覚できることが分かった（非特許文献3を参照されたい。）。このことは、声調的な文脈中に、F0の高さ及びF0の傾斜勾配のほかに、区別のための手がかりが存在することを示している。 20

【0005】

【非特許文献1】D.ワレン及びY.シュウ、「振幅輪郭及び短いセグメントにおける標準中国語の声調情報」、フォネティカ49、1992、pp.25-47。(D. Whalen and Y. Xu, "Information for Mandarin tones in the amplitude contour and in brief segments", *Phonetica* 49, 1992, pp. 25-47.)

【非特許文献2】E.ガーディング他、「現代標準中国語の第四声と第三声との区別」言語と音声、第29巻第3部、1986、pp.281-293。(E. Garding et al., "Tone 4 and tone 3 discrimination in modern standard Chinese", *Language and Speech*, Vol. 29, Part 3, 1986, pp. 281-293.)

【非特許文献3】Y.シュウ、「同時調音声調の生成と知覚」J.A.S.A.(4)、1994、pp.2240-2253。(Y. Xu, "Production and perception of coarticulated tones", *J.A.S.A.* (4), 1994, pp. 2240-2253.) 30

【非特許文献4】Y.R.チャオ、「口語中国語文法」パークレイ、カリフォルニア大学出版社、1968。(Y.-R. Chao, "A grammar of spoken Chinese", Berkeley: University of California Press, 1968.)

【非特許文献5】T.ツムラ他、「周波数遷移の聴覚検出」J.A.S.A.第53巻第1号、1973、pp.17-25。(T. Tsumura et al., "Auditory detection of frequency transition", *J.A.S.A.*, Vol. 53, No. 1, 1973, pp. 17-25.)

【非特許文献6】S.シゲノ、H.フジサキ、「音声及び非音声刺激の範疇判断に対する先行するアンカーの効果」、日本心理学研究、1979、第21巻、第4号、pp.165-173。(S. Shigeno, H. Fujisaki, "Effect of a preceding anchor up on the categorical judgment of speech and non-speech stimuli", *Japanese Psychological Research*, 1979, Vol. 21, No. 4, pp. 165-173.) 40

【非特許文献7】J.S.チャン、K.ヒロセ、「関連付け仮説と中国語連続音声の声調認識へのその応用」ICASSP2000、第III巻、pp.1419-1422。(J.-S. Zhang, K. Hirose, "Anchoring hypothesis and its application to the recognition of Chinese continuous speech", *ICASSP2000*, Vol. III, pp. 1419-1422.)

【非特許文献8】J.S.チャン、S.ナカムラ、K.ヒロセ、「関連付けF0特徴による中国語語彙声調の区別」ICSLP2000予稿集、第II巻、pp.87-90。(J 50

.-S. Zhang, S. Nakamura, and K. Hirose, "Discriminating Chinese lexical tones by anchoring F0 Features", Proc. of ICSLP2000, Vol. II, pp. 87-90.)

【発明の開示】

【発明が解決しようとする課題】

【0006】

これまで、どの研究でも、F0の高さ及びF0の傾斜勾配以外、声調的な文脈中での識別のための手がかりについて明らかに示した研究は存在していない。もしこのような手がかりが利用できれば、中国語等の声調言語で声調を分類する大きな助けとなるはずであり、さらにこのような手がかりによってトレーニングされた適切なモデルがあれば、音声認識又は音声合成装置の性能がさらに改善されるはずである。

10

【0007】

従って、この発明の目的の一つは、中国語の声調の判断を容易にすることである。

【0008】

別の目的は、より自然な品質の中国語音声を作成することである。

【課題を解決するための手段】

【0009】

この発明に従った中国語の声調分類装置は、関連付けにより声調を識別する特徴量を含む特徴量の組が与えられると、中国語で使用される声調分類の確率を出力するように、トレーニングデータセットを用いてトレーニングされた声調モデルを記憶するための手段を含む。関連付けにより声調を識別する特徴量は先頭ギャップ及び末尾ギャップを含む。声調分類装置はさらに、入力された中国語の音声データを一連の声調核にセグメント化するための手段と、声調核の各々から関連付けにより声調を識別する特徴量を抽出するための手段と、抽出する手段によって抽出された音響特徴量を声調モデルに適用し、声調モデルにより出力される最も高い確率を達成する声調分類を選択するためのパターンマッチング手段とを含む。

20

【0010】

好ましくは、関連付けにより声調を識別する特徴量はさらに、声調核の基本周波数(F0)輪郭の傾斜勾配を含む。

【0011】

より好ましくは、関連付けにより声調を識別する特徴量はさらに、声調核の先頭F0及び末尾F0を含む。

30

【0012】

さらに好ましくは、関連付けにより声調を識別する特徴量はさらに、声調核の正規化されたパワーを含む。

【0013】

この発明の別の局面に従った中国語の声調分類装置は、関連付けにより声調を識別する特徴量を含む特徴量の組が与えられると、中国語で使用される声調分類の確率を出力するように、トレーニングデータセットを用いてトレーニングされた声調モデルを記憶するための手段と、入力された中国語の音声データを一連の声調核にセグメント化するための手段と、声調核の各々から関連付けにより声調を識別する特徴量を抽出するための手段とを含む。関連付けにより声調を識別する特徴量は先頭ギャップ及び末尾ギャップを含む。装置はさらに、抽出するための手段によって抽出された先頭ギャップ及び末尾ギャップの符号の組合せに従って声調の分類を判断し、判断された分類を出力するための声調分類手段を含む。

40

【0014】

好ましくは、声調分類手段は、声調の先頭ギャップが正であり末尾ギャップが正であるとき、その声調が中国語の第一声であると判断するための手段を含む。

【0015】

声調分類手段はさらに、声調の先頭ギャップが負であり末尾ギャップが正であるとき、

50

その声調が中国語の第二声であると判断するための手段を含んでも良い。

【 0 0 1 6 】

好ましくは、声調分類手段はさらに、声調の先頭ギャップが負であり末尾ギャップが負であるとき、その声調が中国語の第三声であると判断するための手段を含む。

【 0 0 1 7 】

さらに好ましくは、声調分類手段はさらに、声調の先頭ギャップが正であり末尾ギャップが負であるとき、その声調が中国語の第四声であると判断するための手段を含む。

【 0 0 1 8 】

この発明のさらに別の局面に従った中国語の F 0 生成装置は、関連付けにより声調を識別する特徴量を含む特徴量の組が与えられると、中国語で用いられるそれぞれの声調分類の確率を出力するようにトレーニングデータセットを用いてトレーニングされた声調モデルを記憶するための手段を含む。関連付けにより声調を識別する特徴量は先頭ギャップ及び末尾ギャップを含む。この装置はさらに、構文解析された中国語テキストが与えられると、その中国語テキスト内の音声単位の各々について可能な中国語声調の確率を出力する、確率的 F 0 モデルを記憶するための手段と、確率的 F 0 モデルの出力に従って、入力された中国語のテキストに適合する F 0 のシーケンスを生成するための手段と、生成するための手段によって出力された F 0 シーケンスが声調モデルと整合しているか否かを判断するための手段とを含む。

【発明を実施するための最良の形態】

【 0 0 1 9 】

[前提条件]

(声調的な文脈による F 0 変化現象)

ここで発明者らが検討した文脈による変化は、二つの周知の現象、「階段状の降下」及び「文脈上の同化」である。厳密に言えば、階段状の降下現象は高いピッチターゲットが低いピッチターゲットに続く特定の文脈における、文脈上の同化の一種である。

【 0 0 2 0 】

階段状降下の効果

階段状降下の効果は、「H L H」の声調シーケンスにおいて、「L」が存在するために、二番目の「H」の声調の F 0 の高さが、最初の「H」より低くなる現象として知られている。もし H と L とが交互に現れるシーケンスでこれが連続して作用すると、F 0 の輪郭が階段状の関数となるであろう。ある発話で後の位置にある H の声調の F 0 高さが、同じ発話の前方にある L の声調より低くなることもあり得る。

【 0 0 2 1 】

中国語の四つの基本的語彙声調のうち三つは、図 1 に示すように先頭(第二声)、末尾(第四声)又は先頭と末尾の両方(第三声)のいずれかに低いターゲットを有する。

【 0 0 2 2 】

図 2 は階段状降下現象の例を示す。発話されたテキストは、"you3 qing1 wei1 e4 hua4 xian4 xiang4" (「少しずつ悪化する現象がある。」) である。図 2 で点線 2 0 で示される F 0 シーケンスから、二つの興味深い現象を見ることができる。

【 0 0 2 3 】

1. 階段状降下の効果は後半 3 個の第四声で連続して 3 回起こり、下向きの階段となっている。

【 0 0 2 4 】

2. さらに最後の第四声の H の先頭が発話の始めの第三声の F 0 高さと同じレベルになっている。

【 0 0 2 5 】

語彙的声調の F 0 の高さは階段状降下の影響を相当受けるが、人間のピッチの知覚には何ら干渉しないことが良く知られている。すなわち、図 2 に示す複数の第四声は同じ声調であると知覚されるのみならず、先頭でもほぼ同じ高さであると知覚されるのである。さらに、音声合成の研究では、階段状降下の影響を受けると思われる隣接する二つの声調に

10

20

30

40

50

高いピッチ点で同じ F 0 の値を割り当てると、合成音声では 2 番目の語彙的声調に不自然なストレスがあるように感じられる。

【 0 0 2 6 】

文脈上の同化による声調変化

ここで用いる文脈上の同化による声調変化の概念とは、同化による変化の影響が激しいため、声調の F 0 傾斜の方向が他の声調のものに変わってしまうことさえある、という現象を指す。図 3 は、発話されたテキストが "si1 zhu2 guan3 xian2" (伝統的な弦楽器と木管楽器) である例を示す。図 3 に示した例では、線 3 0 で囲まれた部分の F 0 輪郭は第二声であり、その特徴的なパターンは上昇勾配である。しかしここでは、F 0 の輪郭は平坦なものに変わっており、これは第一声のものである。

10

【 0 0 2 7 】

重要な問題の一つは、F 0 輪郭が変化した場合、声調が別の声調に変わったかどうかである。この問題は非常に分かりにくいので、かつてはかなり相違する見解が示されていた。非特許文献 4 では、図 3 に示すような事例を説明するのに、音響学的な声調連声に関する規則を提案することまでしている。その提案では、先頭が高い声調が先行する第二声であって、その後四つの基本的声調のうちの一つが続く場合、間にはさまれた第二声は、普通の速さの音声では第一声に変化する、と示唆されている。

【 0 0 2 8 】

しかし、最近の研究により、文脈による F 0 変化に続く声調文脈がある限り、聞き手は依然としてその声調を元の声調のものであると知覚することが明らかになった (非特許文献 3)。これは、図 3 における F 0 の変化を受けた第二声が依然として第一声ではなく第二声として知覚される、という意味である。このことは何人かの中国語を母語とする話者によって確認されている。

20

【 0 0 2 9 】

知覚実験の二つの知見

非特許文献 3 は声調文脈による F 0 変化と人間のピッチ知覚に対するその影響とに関し、より体系的な調査を行なっている。非特許文献 3 での興味ある実験の一つは、人間が三音節のシーケンスの中で、間に挟まれた声調をどのように知覚するか、さらに、同じ声調が、その音節をさむ第一音節と第三音節とを交換した文脈では同様に知覚されるか、を検討したものである。第一音節と第三音節とを交換すると、間に挟まれた声調にとっては声調環境がかなり相違することになるので、ピッチ知覚に対する声調文脈の影響を試験するのは特別な方法を提供するものと期待される。

30

【 0 0 3 0 】

図 4 は非特許文献 4 の実験から得られた二つの知見を例示する。

【 0 0 3 1 】

1. 元の声調文脈が自然で交換後の文脈が不自然な場合、間に挟まれた声調は元の文脈と同じ声調であると知覚される傾向がある (図 4 の第 1 行を参照)。

【 0 0 3 2 】

2. 元の声調文脈が不自然で交換後の文脈が自然な場合、間に挟まれた声調は、交換された文脈において元の声調とは逆方向への輪郭を備えた声調であると知覚される傾向がある (図 4 の下の行を参照)。

40

【 0 0 3 3 】

(関連付け仮説及び関連付けによる F 0 特徴量)

語彙的声調を区別するためのより効果的な特徴量を見出すことを目的として、本件発明者らは非特許文献 5 又は非特許文献 6 の心理学的音響学的知覚に関する知見を採用し、以下のような関連付け仮説をたてた (非特許文献 7 及び非特許文献 8 を参照)。

【 0 0 3 4 】

- 最初の語彙的声調の末尾と 2 番目の語彙的声調の先頭との相対的な F 0 の相違は、F 0 輪郭に関する下り勾配という直接的な手がかりの他に、高いピッチ又は低いピッチについて識別するための重要な手がかりとなりうる。

50

【 0 0 3 5 】

- 競合効果について、タイミング割当機構があるはずである（非特許文献6を参照されたい。）。

【 0 0 3 6 】

この仮説に基づき、連続した音声の語彙的声調は、平坦、上昇、途中で下がる、又は下降、というF0パターンを用いること以外に、図5に示された表により示される関連付けによるパターンを用いて、音響的に特徴付けることができる。図5で用いられる用語は以下の通りである。

【 0 0 3 7 】

- 先頭ギャップ：先頭のF0と、先行する語彙の声調の末尾のF0との相違。

10

- 末尾ギャップ：末尾のF0と、後続する語彙の声調の先頭のF0との相違。

【 0 0 3 8 】

（声調文脈による変化の関連付けによる識別）

本件発明者らは、この関連付け仮説を活用して、上述の声調変化をする声調を一貫して予測できることを見出した。

【 0 0 3 9 】

- 階段状に下がる声調の関連付けによる識別

図6は先頭ギャップ及び末尾ギャップの推定方法を例示する。特に、図2の階段状に下がる声調の先頭及び末尾ギャップの推定を例示する。

【 0 0 4 0 】

20

図6において、細い縦線は音節の境界を示す。特徴量“r”は、先頭及び末尾ギャップを表し、特徴量“d”は持続時間を正規化したF0傾斜勾配を表す。先頭及び末尾点は声調核に対応する点である。

【 0 0 4 1 】

第四声は「HL」のピッチパターンを有する。関連付けによる声調識別仮説によれば、以下の条件を満たす場合には第四声と判定される。

【 0 0 4 2 】

1. 先頭ギャップが正である。すなわち $r > 0$ である。第四声は先頭がHであるので、もし先行する声調の末尾がLであれば $r > 0$ となる。一方、先行する声調の末尾がHであれば、 $r < 0$ である。

30

【 0 0 4 3 】

2. 上と同様の機構により、末尾ギャップが負である。

【 0 0 4 4 】

3. F0傾斜の勾配が負、すなわち $d < 0$ （下り勾配のピッチ）である。

【 0 0 4 5 】

図6から、四つの第四声はそれらの絶対的なF0の高さが大きく異なるにも関わらず、上述の条件を満たすという点ではほとんど相違しないことがわかる。四つの第四声のどの第四声も、同じようなレベルの正の先頭ギャップと、同じようなレベルの負の末尾ギャップと、同じようなF0輪郭の下り勾配を有する。これら三つの特徴量はピッチ関連付け仮説によれば、声調を区別するため非常に重要であるので、これら四つの第四声は同様の声調と知覚されるはずであり、リスニングによりそれが実証された。

40

【 0 0 4 6 】

文脈上の同化を受けた声調の関連付けによる区別

図7は第二声に対する先頭及び末尾ギャップrとF0傾斜勾配dの推定を例示する。図7から以下のことがわかる。

【 0 0 4 7 】

1. 先頭ギャップはどちらかといえば負、すなわち $r < 0$ である。

【 0 0 4 8 】

2. 末尾ギャップはどちらかといえば正、すなわち $r > 0$ である。

【 0 0 4 9 】

50

3. F0 輪郭傾斜は平坦、すなわち $d = 0$ である。

【0050】

図5から、上述の特徴量に関して第二声が最も当てはまると分かるので、この声調は第二声であろうと予測する。関連付け仮説に基づく第二声の予測を決定するのは、負の先頭ギャップと正の末尾ギャップである。

【0051】

(交換された文脈での声調の予測)

関連付け声調識別仮説に基づき、上述の交換された文脈での実験で知覚される声調が何かを容易に正確に予測することもできる。図8及び図9はこの予測手順と、交換された文脈で、間に挟まれた声調についての予測結果を示す。ターゲットはまず、先頭ギャップ、
10 末尾ギャップ及び下り勾配のピッチに基づいて高い(+)又は低い(-)ターゲットのいずれであるかが予測され、その後その先頭と末尾との2個のターゲットに基づいてどの声調かが予測可能となる。

【0052】

図8及び図9から、こうして予測された結果が、報告された二つの知見と一貫していることが分かる。

【0053】

中国語の声調区別パターンは音声における声調の音響的特徴量となるパターンを示す。良好なパターンは、発話が分離された音節からなるのか、連続した自然な音声かによらず、
20 実際の音声の音響的特徴量から、一貫した信頼性のある声調の識別を可能とするものであるべきである。

【0054】

関連付けによる声調識別パターンは、連続音声における中国語の声調について提案された最初の識別パターンである。伝統的な中国語の声調パターンと比較して、この新たなパターンは連続音声において声調をより良く識別する能力を提供する。この提案には多くの
応用があるはずである。以下に、これを中国語の自動声調分類と中国語音声の合成に応用したものを簡単に説明する。

【0055】

[第1の実施の形態]

図10はこの発明の第1の実施の形態に従った声調分類システム100を例示する。システム100は上述の関連付けによる声調識別特徴量に基づいたものである。
30

【0056】

図10を参照して、システム100は、一組の関連付けによる特徴量が与えられると最も確率の高い声調分類を出力するように声調モデル106をトレーニングデータ102を用いてトレーニングするための声調モデルトレーニングユニット104と、入力された音声108に
応答して、声調モデルトレーニングユニット104によってトレーニングされた声調モデル106を用いて、声調分類112を出力する声調分類ユニット110とを含む。

【0057】

声調モデル106はそれぞれの声調に対して準備された多数のガウス混合モデル(GMM)を含む。文脈に依存しないGMMのみを用いる場合、GMMの総数は5である。すな
40 わち、基本の四声について4個と、軽声について1個である。文脈に依存するGMMを用いるのであれば、最大で175個のGMMを使用可能である(三つの声調からなる文脈に $5 * 5 * 5 = 125$ 、左境界の声調に $4 * 5$ 、右境界の声調に $5 * 5$ 、孤立した声調に5である)。

【0058】

声調モデルトレーニングユニット104は、トレーニングデータ102の発話の各々の声調核を音響的にセグメント化するための音響的セグメント化モジュール130と、音響的セグメント化モジュール130によってセグメント化された音声信号の声調核のF0及び
50 パワーを含む音響的特徴量を抽出するための特徴抽出モジュール132と、特徴抽出モ

ジュール 1 3 2 によって抽出された音響的特徴量から、先頭ギャップ、末尾ギャップ、F 0 輪郭勾配等の関連付けによる特徴量を抽出するための関連付けによる特徴抽出モジュール 1 3 4 と、関連付けによる特徴抽出モジュール 1 3 4 によって抽出された関連付け特徴量を利用して、声調モデル 1 0 6 のモデルパラメータを推定するための声調モデル推定モジュール 1 3 6 とを含む。

【 0 0 5 9 】

音響的セグメント化モジュール 1 3 0 での声調核の音響的セグメント化は、統計的方法により、又は音声認識装置の音声セグメント化を利用するだけで実現できる。

【 0 0 6 0 】

声調分類ユニット 1 1 0 は、入力された音声 1 0 8 をセグメント化し、セグメント化された音声信号の声調核を出力するための音響的セグメント化モジュール 1 5 0 と、音響的セグメント化モジュール 1 5 0 によってセグメント化された音声信号の声調核の F 0 及びパワーを含む音響的特徴量を抽出するための特徴抽出モジュール 1 5 2 と、特徴抽出モジュール 1 5 2 によって抽出された音響的特徴量から関連付けによる特徴量を抽出するための関連付けによる特徴抽出モジュール 1 5 4 と、声調モデル 1 0 6 のうち関連付けによる特徴抽出モジュール 1 5 4 によって抽出された関連付けによる特徴量に最も良く一致する特徴パターンを探索し、最も良く一致するパターンに対応する声調分類 1 1 2 を出力するためのパターンマッチングモジュール 1 5 6 とを含む。

10

【 0 0 6 1 】

システム 1 0 0 は以下のように動作する。このシステムには二つの動作局面がある。トレーニングの局面と動作の局面とである。

20

【 0 0 6 2 】

トレーニングの局面では、トレーニングデータ 1 0 2 内の発話が、音響的セグメント化モジュール 1 3 0 によって声調の境界で音響的にセグメント化される。セグメント化された音声信号の声調核が特徴抽出モジュール 1 3 2 に与えられ、F 0 及びパワーを含むセグメント化された音声信号の音響的特徴量が抽出される。

【 0 0 6 3 】

関連付けによる特徴抽出モジュール 1 3 4 は声調核の音響的特徴量から、関連付けに基づく声調識別のための特徴量を抽出する。二つの最も重要な関連付けによる識別特徴量は先頭ギャップと末尾ギャップであり、これらは図 1 1 により、 $r_{1,1}$ 及び $r_{2,2}$ として計算される。別の重要な特徴量は F 0 輪郭の傾斜勾配、先頭 F 0、末尾 F 0 である。必要に応じて、声調核の正規化されたパワーを含めても良い。この場合、1 個の声調の特徴量は 5 次元又は 6 次元のベクトルとなる。

30

【 0 0 6 4 】

声調モデルトレーニング処理により、各声調についての声調モデル 1 0 6 内のガウス混合モデルのパラメータがトレーニングデータ 1 0 2 で推定される。

【 0 0 6 5 】

次の、声調認識 / 分類局面では、入力された音声 1 0 8 が音響的セグメント化モジュール 1 5 0 でセグメント化され、特徴抽出モジュール 1 5 2 で、セグメント化された音声信号から音響的特徴量が抽出される。さらに、特徴抽出モジュール 1 5 2 によって抽出された特徴量から、関連付けによる特徴抽出モジュール 1 5 4 により、関連付けによる特徴量が抽出される。

40

【 0 0 6 6 】

関連付けによる特徴量のパターンが声調モデルと比較される。最大確率を有する声調のシーケンスが、認識された声調のシーケンスとして出力される(声調分類 1 1 2)。

【 0 0 6 7 】

[第 2 の実施の形態]

関連付けによるパターンは、従来の規則ベースの合成システムのようにテキストから F 0 パターンを生成するために直接利用するばかりでなく、確率的 F 0 生成システムと統合して、合成音声の声調的分かりやすさを改善することもできる。

50

【 0 0 6 8 】

規則ベースのイントネーション生成システムでは、トレーニングコーパスから開発した多数の規則を用いて合成すべきテキストのイントネーションF0パターンを予測する。関連付け声調識別パターンをこの規則の組と統合して、声調を識別することが十分可能となる手がかりを、生成されたF0から得られるようにすることが容易にできる。言い換えれば、標準的な声調パターンと比較して、関連付けによるパターンはF0の軌跡の配置に対し自由度が高い。これにより、合成音声の自然さが改善されるはずである。

【 0 0 6 9 】

現在の確率的F0生成システムでは、合成処理に先だってトレーニング期間を設ける。このトレーニング処理で特徴的な事項は、関連する文脈に関するファクタの数が多すぎて、頑健なイントネーションモデルを構築するために十分なトレーニングデータが得られない、ということである。

10

【 0 0 7 0 】

中国語を例にとると、約1300の声調音節があり、可能な三音節のつながりは22億を超える。強調や韻律句内での位置（始点、中位、終点）、疑問文か平叙文か、等のさらなるファクタを考慮するとすれば、可能な組合せは兆の単位になりうる。このような膨大な文脈的な組合せについてモデルをトレーニングするに足るデータを収集することは不可能である。従って、深刻なデータスパースネスの問題により、生成されたイントネーションのF0が奇妙に聞こえることは多々ある。

【 0 0 7 1 】

しかし、関連付けに基づく声調区別パターンを補助として用いることにより、確率的に合成されたF0輪郭を検討するための事後チェックモジュールを簡単に利用することができる。

20

【 0 0 7 2 】

図12は、入力されたテキスト200からF0シーケンス204を生成するための、この発明の第2の実施の形態によるF0生成ユニット202を示す。F0生成ユニット202は先行技術のF0生成装置で用いられる確率的F0モデル206、及びこの発明の第1の実施の形態で用いた声調モデル208と共に用いられる。

【 0 0 7 3 】

F0生成ユニット202は、入力されたテキスト200を構文解析するためのテキスト構文解析モジュール220と、確率的F0モデル206を用いてテキスト構文解析モジュール220によって構文解析されたテキストのイントネーションF0を生成するためのイントネーションF0生成モジュール222と、声調モデル208を参照してイントネーションF0生成モジュール222が出力したF0シーケンスをチェックし、パターンに合わないF0を訂正して、訂正されたF0シーケンス204を出力する事後チェックモジュール224とを含む。

30

【 0 0 7 4 】

上述の通り、関連付けによる特徴パターンは、話す速度の変化、感情の変化等の高いレベルでの話し方のスタイルに従って、合成されたF0輪郭を修正する際に、確実に声調を識別可能にする。この問題は、先行技術の中国語TTS(テキスト 音声合成)システムではほとんど実現され得ないものである。

40

【 0 0 7 5 】

上述の実施の形態は単なる例示であって制限的なものと解してはならない。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及び範囲内でのすべての変更を含む。

【 図面の簡単な説明 】

【 0 0 7 6 】

【 図 1 】 中国語の声調パターンを示す図である。

【 図 2 】 4個の連続した第四声で観察された階段状に下降する効果の例を示す図である。

【 図 3 】 文脈上の同化の例を示す図である。

50

【図4】間に挟まれた語彙声調の声調知覚に対する、文脈交換の影響を示す図である。

【図5】連続音声における4個の基本的語彙声調の、関連付けによる特徴量を示す図である。

【図6】図2の階段状に下がる声調について、先頭ギャップと末尾ギャップとの推定を示す図である。

【図7】図3の2番目の声調について、先頭ギャップと末尾ギャップとの推定を示す図である。

【図8】交換文脈の知覚実験において、声調変化の予測を示す図である。

【図9】交換文脈の知覚実験において、声調変化の予測を示す図である。

【図10】この発明の第1の実施の形態に従った声調分類システム100を示すブロック図である。

10

【図11】関連付けによる声調区別特徴量の抽出を示す図である。

【図12】声調モデルの関連付けによる声調パターンによりチェックされ訂正されたF0シーケンスを生成するための、この発明の第2の実施の形態に従ったF0生成ユニット202を示す図である。

【符号の説明】

【0077】

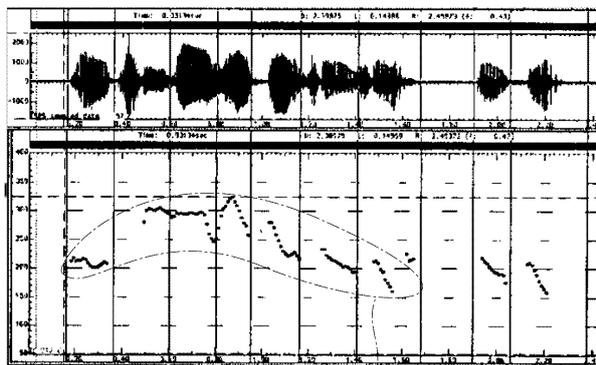
100 声調分類システム、102 トレーニングデータ、104 声調モデルトレーニングユニット、106 声調モデル、108 入力音声、110 声調分類ユニット、112 声調分類、130, 150 音響的セグメント化モジュール、132, 152 特徴抽出モジュール、134, 154 関連付けによる特徴抽出モジュール、136 声調モデル推定モジュール、156 パターンマッチングモジュール、200 入力テキスト、202 F0生成ユニット、204 F0シーケンス、206 確率的F0モデル、208 声調モデル、220 テキスト構文解析モジュール、222 イントネーションF0生成モジュール、224 事後チェックモジュール

20

【図1】

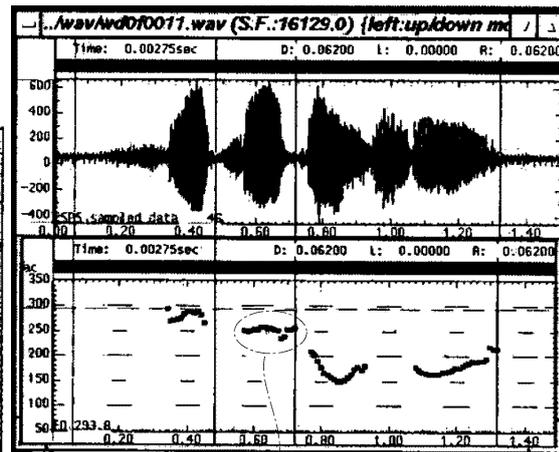
ターゲット	第一声	第二声	第三声	第四声
先頭	H	L	L	H
末尾	H	H	L	L

【図2】



20

【図3】



30

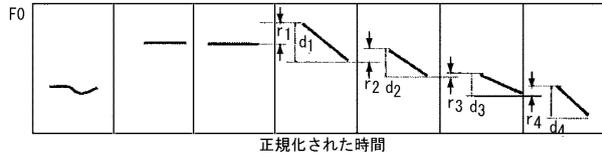
【図4】

文脈変化	基になる形	同時調音後	交換後	知覚結果
自然から不自然へ				
不自然から自然へ				

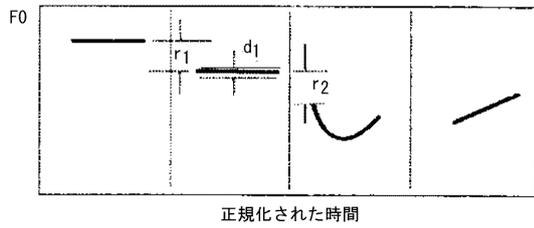
【図5】

語彙声調	先頭ギャップ	末尾ギャップ
第一声	≥ 0	≥ 0
第二声	≤ 0	≥ 0
第三声	≤ 0	≤ 0
第四声	≥ 0	≤ 0

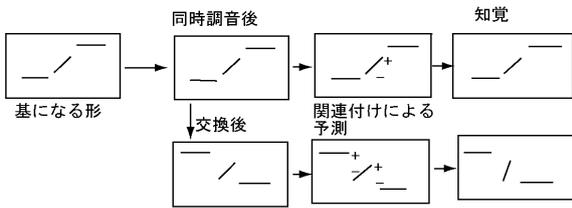
【図6】



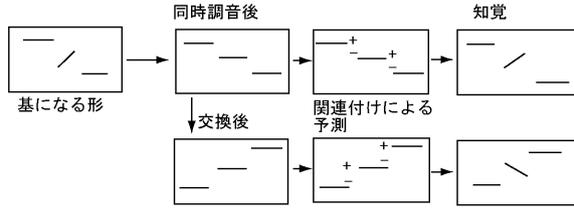
【図7】



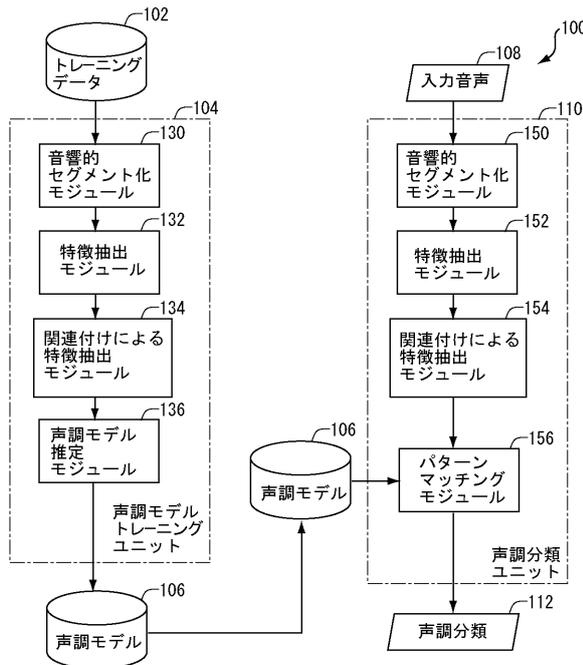
【図8】



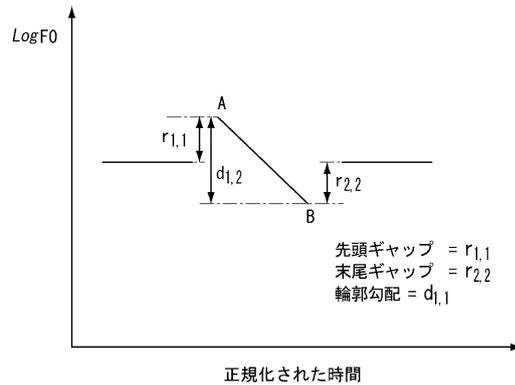
【図9】



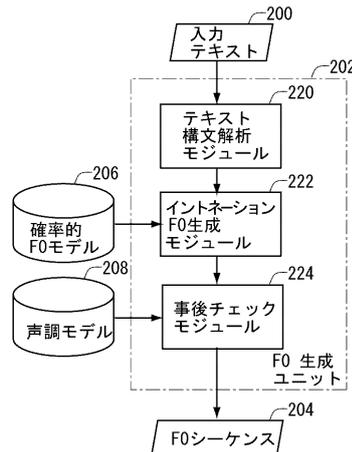
【図10】



【図11】



【図12】



フロントページの続き

(56)参考文献 特開2003-058182(JP,A)

特開平08-263097(JP,A)

特開2003-330482(JP,A)

Jin-Song Zhang, Satoshi Nakamura and Keikichi Hirose, Discriminating Chinese Lexical Tones by Anchoring F0 Features, 6th International Conference on Spoken Language Processing(ICSLP 2000), ISCA, 2000年10月16日, Vol.2, p.87-90

Jin-Song Zhang and Keikichi Hirose, Anchoring Hypothesis and Its Application to Tone Recognition of Chinese Continuous Speech, IEEE International Conference on Acoustics, Speech, and Signal Processing, 米国, IEEE, 2000年6月5日, 2000 Vol.3, p.1419-1422

胡力游、今井 聖, 中国語連続音声の声調認識, 電子情報通信学会論文誌・A 基礎・境界, 日本, 社団法人電子情報通信学会, 1995年7月, J78-A(7), p.763-768

(58)調査した分野(Int.Cl., DB名)

G10L 15/00 - 15/28

G10L 13/00 - 13/08