

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2010-26068

(P2010-26068A)

(43) 公開日 平成22年2月4日(2010.2.4)

|                                |                       |             |
|--------------------------------|-----------------------|-------------|
| (51) Int.Cl.                   | F I                   | テーマコード (参考) |
| <b>G 1 O L 15/10 (2006.01)</b> | G 1 O L 15/10 5 O O N | 5 D O 1 5   |
| <b>G 1 O L 11/00 (2006.01)</b> | G 1 O L 11/00 4 O 2 H |             |
|                                | G 1 O L 11/00 1 O 1 C |             |

審査請求 未請求 請求項の数 6 O L (全 9 頁)

|           |                              |          |                     |
|-----------|------------------------------|----------|---------------------|
| (21) 出願番号 | 特願2008-184897 (P2008-184897) | (71) 出願人 | 000003078           |
| (22) 出願日  | 平成20年7月16日 (2008.7.16)       |          | 株式会社東芝              |
|           |                              |          | 東京都港区芝浦一丁目1番1号      |
|           |                              | (74) 代理人 | 100109900           |
|           |                              |          | 弁理士 堀口 浩            |
|           |                              | (72) 発明者 | 杉浦 千加志              |
|           |                              |          | 東京都港区芝浦一丁目1番1号 株式会社 |
|           |                              |          | 東芝内                 |
|           |                              | Fターム(参考) | 5D015 CC07 CC13     |

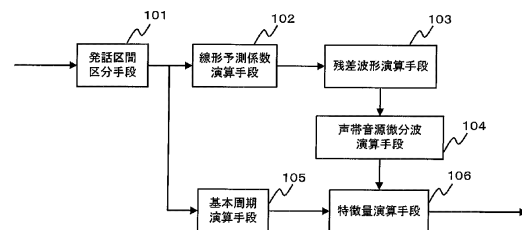
(54) 【発明の名称】 情報処理装置

(57) 【要約】

【課題】 学習を必要とせずに感情抽出に必要な声の張りに関する特徴量の抽出を行う技術を提供する。

【解決手段】 有声音において、入力音声の声帯音源微分波を推定し、この声帯音源微分波から基本周期ごとに現れる急峻な立ち上がりの最小位置と最大位置を特定し、立ち上がりの最大位置から次の立ち上がりの最小位置までの区間の声帯音源微分波が有する特徴を数値化することによって基本周期やパワーに依存することなく、声の張りに関する特徴量を算出する。

【選択図】 図2



**【特許請求の範囲】****【請求項 1】**

音声を入力する入力手段と、

前記入力手段によって入力された音声をフレームに区分する区分手段と、

前記区分手段から出力されたフレーム音声に対して線形予測分析を行って、線形予測の残差波形を算出し、この残差波形に基づいて声帯音源波に放射特性を加えた声帯音源微分波を推定する声帯音源微分波演算手段と、

前記区分手段によって区分されるフレーム音声から基本周期を算出する基本周期演算手段と、

前記声帯音源微分波演算手段によって推定された声帯音源微分波に対して、前記基本周期ごとに急峻な立ち上がり位置を特定し、隣り合う急峻な立ち上がり位置間の前記声帯音源微分波から特徴量を算出する特徴量算出手段とを有することを特徴とする情報処理装置。

10

**【請求項 2】**

前記特徴量算出手段は、前記声帯音源微分波演算手段によって推定された声帯音源微分波に対して、前記基本周期ごとに急峻な立ち上がりの最小位置と最大位置とを特定して、急峻な立ち上がりの最大位置と次の急峻な立ち上がりの最小位置とを結ぶ線分を取得し、前記声帯音源微分波と前記線分との差分の合計値を特徴量として算出することを特徴とする請求項 1 に記載の情報処理装置。

**【請求項 3】**

前記特徴量算出手段は、前記声帯音源微分波演算手段によって推定された声帯音源微分波に対して、前記基本周期ごとに急峻な立ち上がりの最小位置と最大位置とを特定して、急峻な立ち上がりの最大位置と次の急峻な立ち上がりの最小位置とを結ぶ線分を取得し、前記声帯音源微分波と前記線分との差分の荷重和を特徴量として算出することを特徴とする請求項 1 に記載の情報処理装置。

20

**【請求項 4】**

前記基本周期演算手段によってフレーム音声の基本周期が算出されなかった場合、当該フレームに対して前記特徴量算出手段による特徴量の算出を行わないことを特徴とする請求項 1 に記載の情報処理装置。

**【請求項 5】**

前記声帯音源微分波の前記基本周期ごとの急峻な波形の立ち上がり位置は、前記基本周期ごとに声帯音源微分波を微分して得られる値が最大値のサンプルから前後に前記微分して得られる値が前記最大値の所定の割合以上のサンプルが連続する区間とすることを特徴とする請求項 1 に記載の情報処理装置。

30

**【請求項 6】**

入力された音声をフレームに区分する区分機能と、

前記区分機能によって得られるフレーム音声に対して線形予測分析を行って、線形予測の残差波形を算出し、この残差波形に基づいて声帯音源波に放射特性を加えた声帯音源微分波を推定する声帯音源微分波演算機能と、

前記区分機能によって区分されるフレーム音声から基本周期を算出する基本周期演算機能と、

40

前記声帯音源微分波演算機能によって推定された声帯音源微分波に対して、前記基本周期ごとに急峻な立ち上がり位置を特定し、隣り合う急峻な立ち上がり位置間の前記声帯音源微分波から特徴量を算出する特徴量算出機能とを有することを特徴とする音声特徴量算出プログラム。

**【発明の詳細な説明】****【技術分野】****【0001】**

本発明は情報処理装置に関する。

**【背景技術】**

50

## 10

10

## 10

10

10

## 10

## 10

## 10

10

## 10

10

## 10

10

## 10

## 10

10

## 10

## 10

10

## 【発明を実施するための最良の形態】

## 【0009】

以下、本発明の実施形態について図面を参照して説明する。

## 【0010】

図1は、本実施形態に係る情報処理装置の構成を示すブロック図である。情報処理装置1は、制御部51、記憶部52、操作部53、音声入出力部54、表示部55を備えている。制御部51はCPU、ROM、RAMなどによって構成され、記憶部52に記憶されたプログラムを実行し、情報処理装置1全体を制御する。その1つとして制御部51は、声帯特徴量取得部100を有する。操作部53は、ユーザからの操作を制御部51へ伝える。音声入出力部54は、スピーカやマイクを含み、マイクによって入力された音を制御部51へ伝える。また、制御部51の制御に従って、スピーカから音を出力する。表示部55は、液晶ディスプレイなどによって構成され、制御部51の制御に従って文字や画像を表示する。

10

## 【0011】

声帯特徴量取得部100は、例えば音声入出力部54のマイクを用いて入力された音声や、記憶部52に記憶されている音声などが入力され、声の張りを表す値（以降、声帯特徴量と称する）を出力する。この声の張りとは、声帯の振動によって決まると考えられる。声帯が振動することによって起こる声帯音源波は、声道を通して、口から放射されることで音声として発せられる。この放射による周波数特性（放射特性）は微分によって近似されることが一般的に知られている。つまり、音声波形Sは、声帯音源波と放射特性を表すために声帯音源波を微分して得られる波形（以降、声帯音源微分波と称する）Gと、声道特性Vとを用いて、 $S = G \cdot V$ と表すことができる。そこで、声帯特徴量取得部100では、入力される音声波形Sから声帯音源微分波Gを推定し、推定した声帯音源微分波Gを用いて、声帯特徴量を算出して出力する。

20

## 【0012】

図2は、声帯特徴量取得部100の機能を示すブロック図である。声帯特徴量取得部100は、発話区間区分手段101、線形予測係数演算手段102、残差波形演算手段103、声帯音源微分波演算手段104、基本周期演算手段105、声帯特徴量演算手段106を有する。

## 【0013】

30

声帯特徴量取得部100には、例えば音声入出力部54のマイクを用いて入力された音声や、記憶部52に記憶されている音声などが入力される。入力音声は、発話区間区分手段101に入力される。

## 【0014】

発話区間区分手段101では、入力された音声を所定の時間長（例えば30msec）のフレームに区分する。発話区間区分手段によって区分されたフレーム音声は、線形予測係数演算手段102に入力するとともに、基本周期演算手段105に入力する。

## 【0015】

線形予測係数演算手段102では、入力されたフレーム音声に対して窓掛けを行ったうえで線形予測分析を行い、線形予測係数を算出する。算出した線形予測係数は、残差波形演算手段103に出力する。

40

## 【0016】

残差波形演算手段103では、線形予測係数を用いて声道逆フィルタを形成し、その声道逆フィルタに線形予測係数演算手段102で窓掛けを行ったフレーム音声を入力して、残差波形を取得し、声帯音源微分波演算手段104へ出力する。線形予測分析によって算出された線形予測係数は主に声道特性に関する情報を有しているため、残差波形はおおよそ音声の声道特性を除去した波形とみなすことができる。この残差波形を用いて後述の声帯特徴量の算出処理を行うため、声道特性（個人差）が声帯特徴量に及ぼす影響を低減することができる。

## 【0017】

50

声帯音源微分波演算手段 104 では、残差波形を積分して声帯音源微分波を推定する。残差波形と抽出されるべき声帯音源微分波とを比較すると、残差波形はフラットな周波数特性を有するのに対して、声帯音源微分波は、一般的に  $-6 \text{ dB/oct}$  の傾斜（周波数が 2 倍になる（1 オクターブ上がる）ごとに  $6 \text{ dB}$  下がる）を持つことが知られている。線形予測分析は、残差波形の周波数特性がフラットになるように線形予測係数を算出する処理なので、残差波形からは、声道特性のみならず、声帯音源微分波が本来有する周波数特性も失われていることになる。そこで、声帯音源微分波を得るため、残差波形に対して  $-6 \text{ dB/oct}$ （1 オクターブ上がるごとに  $6 \text{ dB}$  下げるような）処理を行う。この  $-6 \text{ dB/oct}$  の処理には、例えば積分処理を用いる。離散信号に対する積分処理は、入力信号を  $x(n)$ 、積分処理後の出力信号を  $y(n)$ 、係数  $\alpha$  を  $0 < \alpha < 1$  とすると、式 (1) によって実現することができる。

10

【0018】

$$y(n) = x(n) + \alpha \cdot y(n-1) \quad (1)$$

このような積分処理を行うと、 $\alpha$  が 1.0 に近い場合、 $-6 \text{ dB/oct}$  の特性が付加される。残差波形はフラットな周波数特性を有するため、積分処理を行うことで、 $-6 \text{ dB/oct}$  の特性を付加することができる。なお、残差波形から声帯音源微分波を得る処理は積分処理に限定することではなく、例えば  $-6 \text{ dB/oct}$  の特性を有する FIR フィルタによって処理しても良い。

【0019】

以上の発話区間区分手段 101、線形予測係数演算手段 102、残差波形演算手段 103、声帯音源微分波演算手段 104 の処理によって声帯音源微分波を推定することができる。

20

【0020】

次に、図 3 を参照して、本発明が強い声と弱い声の程度の数値化（判別）に有効である理由を説明する。強い声と弱い声の「本来抽出されるべき声帯音源微分波」と「抽出される声帯音源微分波」を比較する。図 3 は、強い声と弱い声それぞれの本来抽出されるべき声帯音源微分波と抽出される声帯音源微分波の一例であって、「本来抽出されるべき声帯音源微分波」は LF モデルと呼ばれるパラメトリックな声帯音源波系モデルで生成しており、「抽出される声帯音源微分波」は、この LF モデルで生成した声帯音源微分波（本来抽出されるべき声帯音源微分波）を駆動源とし、全極モデルを声道フィルタとする、ソース

30

【0021】

図 3 の例では、強い声の場合は、本来抽出されるべき声帯音源微分波の波形と抽出される声帯音源微分波の波形とが大差ないことが分かる。それに対して、弱い声は、本来抽出されるべき声帯音源微分波の波形と抽出される声帯音源微分波の波形との間に大きな差があることがわかる。このように、抽出される声帯音源微分波は、強い声か弱い声かに応じた特徴が表れる。

【0022】

そこで、特徴量演算手段 106 では、声帯音源微分波演算手段 104 から入力される声帯音源微分波が、声の強さに応じて異なる特徴を有することを利用して、声の強さを表す特徴量を算出する。この特徴量の算出は、基本周期演算手段 105 によって算出される基本周期を用いて処理を行う。

40

【0023】

基本周期演算手段 105 で基本周期を算出する場合には、発話区間区分手段 101 から出力されるフレームに切り分けられた入力音声から抽出しても良いし、残差波形演算手段 103 から出力される線形予測残差信号から抽出しても良いし、線形予測残差信号を低域濾波した信号を用いても良い。（図 2 のブロック図では、発話区間区分手段 101 から出力されるフレーム音声の基本周期演算手段 105 に入力される場合を示している。）また、基本周期抽出の手法も例えば自己相関を用いて自己相関係数のピーク位置を基本周期として抽出する手法などが考えられるが、特に限定されるものではない。なお、基本周期演

50

算手段 105 によって基本周期が算出できない場合には、無声音であると判断して、特徴量演算手段 106 に対して当該フレーム音声が無声音であることを示す情報を出力する。この情報を特徴量演算手段 106 が受けると、当該フレームに対する特徴量算出の処理は行わない。基本周期演算手段 105 によって基本周期が算出できない場合とは、例えば自己相関係数のピーク位置を用いて基本周期を取得する際、自己相関係数のピーク位置が人の声の基本周期の範囲から逸脱した箇所に現れたり、ノイズのようなピークが得られたりする場合である。

#### 【0024】

特徴量演算手段 106 には、声帯音源微分波演算手段 104 から出力されるフレームごとの声帯音源微分波と、基本周期演算手段 105 から出力されるフレームごとの基本周期とが入力される。特徴量演算手段 106 は、声帯音源微分波が基本周期ごとに有する急峻な立ち上がり位置を特定し、隣り合う立ち上がり位置で挟まれた区間の波形によって強い声か弱い声かを表す特徴量を算出する。

#### 【0025】

声帯音源微分波を用いて強い声か弱い声かを表す特徴量の算出の概要について図 4 を用いて説明する。図 4 は、強い声と弱い声の特徴的な声帯音源微分波の例を表している。強い声の場合、ある立ち上がりの最大位置 2 T から次の立ち上がりの最小位置 3 B までを結んだ線分よりも声帯音源微分波の各サンプルの方が高い値となる。それに対して、弱い声の場合、ある立ち上がりの最大位置 2 T から次の立ち上がりの最小位置 3 B までを結んだ線分よりも声帯音源微分波のサンプルの方が低い値となる場合が多い。そこで、ある立ち上がり位置の最大位置から次の立ち上がり位置の最小位置の区間において、声帯音源微分波から立ち上がり位置によって特定される線分を差し引いた差分の合計値を考える。この差分の合計値は、図 4 のハッチングで示した範囲の面積を求めることに相当する。ただし、立ち上がり位置によって特定される線分よりも上に位置するハッチングで示した範囲の面積はプラスの値であるのに対して、線分よりも下に位置するハッチングで示した範囲の面積はマイナスの値である。

#### 【0026】

また、総和のみならず、重み付き総和を算出しても良い。予め、結果が良くなるような重みを用意しておき、この重み係数を用いて重み付き総和を算出することで、声の強さの数値化精度を向上させることができる。

#### 【0027】

具体的には、主観などにより声の張り度合いを点数付けした音声を用意しておき、この音声から声の張りに関する特徴量を算出し、この点数と特徴量との相関がなるべく大きくなるように重み係数を算出することで、高精度な声の張りに関する特徴量の抽出が可能となる。重み係数の算出方法は、例えばヒューリスティックに行っても良いし、GA（遺伝的アルゴリズム）や NN（ニューラルネットワーク）などの逐次更新を繰り返す手法を用いても良い。また図 5 の様に、あるフレーム n において、線分が位置する区間を区間数 M に分割し、この小区間ごとの線分と声帯音源微分波で囲まれる小面積を  $S_{nm}$  とし、重み係数を  $w_m$  とすると、そのフレームにおける特徴量  $X_n$  は式 (2) で表されるので、この  $X_n$  と主観評価点数  $Y_n$  との相関を最大化するために、正準相関分析を用いても良い。こうすることで、声の張りに関する特徴量と主観などで点数付けした声の張り度合いとの相関を最大化するための重み係数を解析的に算出することができ、結果的に声の張りに関する特徴量の抽出精度を向上することができる。

#### 【数 1】

$$X_n = \sum_{m=1}^M (w_m \cdot S_{nm}) \quad (2)$$

#### 【0028】

図 4 の場合、強い声の声帯音源微分波では、ハッチングで示した範囲は線分よりも上に

位置しているため、差分の合計値がプラスの値となる。それに対して弱い声の声帯音源波では、ハッチングで示した範囲は線分よりも下に位置しているのがほとんどであるため、差分の合計値がマイナスの値となる。

#### 【 0 0 2 9 】

このように、声帯音源微分波と立ち上がり位置によって特定される線分との差分は、強い声と弱い声によって大きく異なるため、この差分に基づいて強い声か弱い声かを定量的に表す特徴量を算出することができる。

#### 【 0 0 3 0 】

図 6 は特徴量演算手段 1 0 6 の具体的な処理を示すフローチャートである。まず特徴量演算手段 1 0 6 は、入力されたフレームごとの声帯音源微分波を微分する ( S 2 0 1 ) 。  
なお、声帯音源微分波演算手段 1 0 4 において、残差波形を積分することによって声帯音源微分波を算出する場合、残差波形が声帯音源微分波を微分した波形に相当するため、残差波形演算手段 1 0 3 の出力である残差波形を特徴量演算部 1 0 6 に入力して、声帯音源微分波を微分した波形として用いても良い。

10

#### 【 0 0 3 1 】

次に、基本周期ごとに声帯音源微分波を微分して得られた値が最大のサンプルを探索する ( S 2 0 2 ) 。このサンプルが声帯音源微分波の増加量が最大になる位置である。そして、最大のサンプルから前後に声帯音源微分波を微分して得られた値が最大値の所定の割合以上の値を有するサンプルがどこまで連続しているかを探索する。この所定の割合は、経験的に求められる値であって、ここでは例えば 7 0 % であるとして説明する。ある基本周期内で声帯音源微分波を微分して得られた値が最大のサンプルを基準として、前方向に 1 サンプルずつ探索し、声帯音源微分波を微分して得られた値が当該最大値の 7 0 % を維持している連続したサンプルのうち、一番前のサンプルを見つけ、この一番前のサンプルを急峻な立ち上がりの最小位置とする ( S 2 0 3 ) 。また、後ろ方向も同様に 1 サンプルずつ探索することによって、最大値の 7 0 % を維持している連続した一番後ろのサンプルを見つけ、この一番後ろのサンプルを急峻な立ち上がりの最大位置とする ( S 2 0 4 ) 。

20

#### 【 0 0 3 2 】

そして、急峻な立ち上がりの最小位置と 1 つ前の急峻な立ち上がりの最大位置とを結ぶ線分を特定し ( S 2 0 5 ) 、この 1 つ前の立ち上がりの最大位置と次の立ち上がりの最小位置の区間で、声帯音源微分波と線分との差分の合計値を算出する ( S 2 0 6 ) 。そして、この合計値を基本周期で除算した結果を特徴量として出力する ( S 2 0 7 ) 。基本周期は、個人差や、言語や方言によって異なる、抑揚によって変化するため、合計値を基本周期で除することによって、特徴量から基本周期の影響を低減することができる。

30

#### 【 0 0 3 3 】

なお、基本周期ごとの声帯音源微分波と線分との差分の合計値を基本周期で除した値を特徴量として出力するとしたが、基本周期ごとの声帯音源微分波と線分との差分の合計値を基本周期で除した値をフレームなど所定の単位ごとに合計した値を特徴量として出力しても良い。

#### 【 0 0 3 4 】

また、特徴量を算出するときに、基本周期で除するのではなく、フレーム単位の実効値や基本周期単位の実効値で除しても良いし、最大振幅で除しても良い。このような値で除することによって、声の大きさ、マイクの距離、マイクゲインなどパワーの違いの特徴量への影響を低減することができる。

40

#### 【 0 0 3 5 】

以上のような実施形態の構成をとることで、基本周期やパワーに依存しない特徴を抽出するため、話し方や声の大きさの影響を低減して、感情抽出に必要な声の張り ( 声の強さ ) を表す特徴量を高精度に抽出することができる。また、声の張りを表す特徴量を抽出するときには、学習を必要としないため、低演算量かつ特別なチューニングを必要とせずに特徴量の抽出をおこなうことができる。

#### 【 0 0 3 6 】

50

なお、上記実施形態に限定されることはなく、本発明の要旨を逸脱しない範囲において、適宜変更しても良い。

【図面の簡単な説明】

【0037】

【図1】本発明の実施形態に係る情報処理装置の構成を示すブロック図。

【図2】本発明の実施形態に係る声帯特徴量取得部の機能を示すブロック図。

【図3】強い声から抽出される声帯音源微分波と弱い声から抽出される声帯音源微分波の違いを示す図。

【図4】本発明の実施形態に係る情報処理装置の特徴量演算手段での処理を説明するための図。

【図5】本発明の実施形態に係る情報処理装置の特徴量演算手段での処理を説明するための図。

【図6】本発明の実施形態に係る情報処理装置の特徴量演算手段の処理フローチャート。

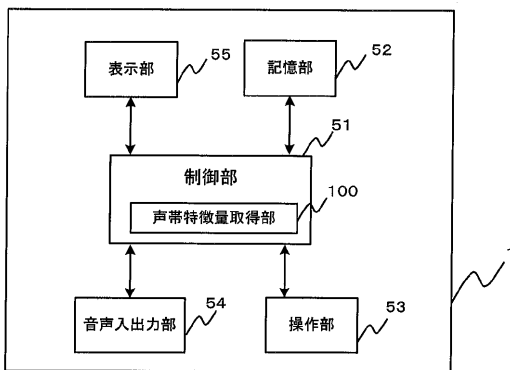
【符号の説明】

【0038】

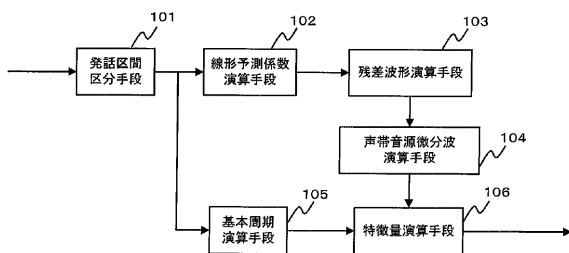
1 情報処理装置、51 制御部、52 記憶部、53 操作部、54 音声入出力部、55 表示部、100 声帯特徴量取得部、101 発話区間区分手段、102 線形予測係数演算手段、103 残差波形演算手段、104 声帯音源微分波演算手段、105 基本周期演算手段、106 声帯特徴量演算手段

10

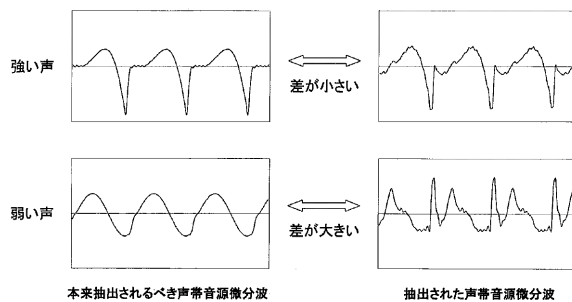
【図1】



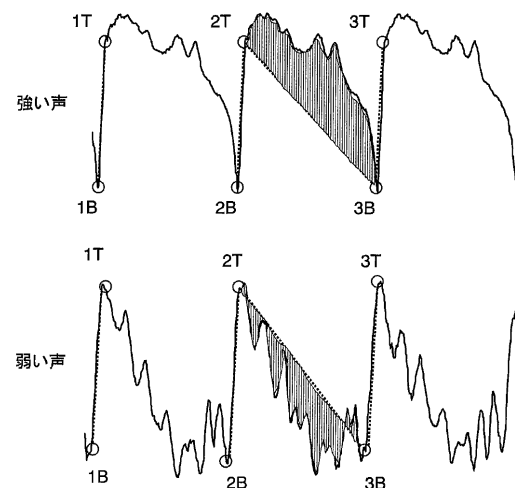
【図2】



【図3】

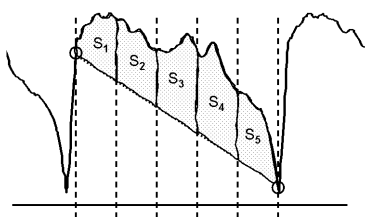


【図4】



【図 5】

M=5の場合



【図 6】

