

(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2019-197149
(P2019-197149A)

(43) 公開日 令和1年11月14日(2019. 11. 14)

(51) Int. Cl.	F I	テーマコード (参考)
G 1 O L 21/0364 (2013.01)	G 1 O L 21/0364	
G 1 O L 19/26 (2013.01)	G 1 O L 19/26	B
G 1 O L 21/0332 (2013.01)	G 1 O L 21/0332	

審査請求 未請求 請求項の数 5 O L (全 28 頁)

(21) 出願番号	特願2018-91200 (P2018-91200)	(71) 出願人	000004226 日本電信電話株式会社 東京都千代田区大手町一丁目5番1号
(22) 出願日	平成30年5月10日 (2018. 5. 10)	(74) 代理人	100121706 弁理士 中尾 直樹
		(74) 代理人	100128705 弁理士 中村 幸雄
		(74) 代理人	100147773 弁理士 義村 宗洋
		(72) 発明者	鎌本 優 東京都千代田区大手町一丁目5番1号 日 本電信電話株式会社内
		(72) 発明者	杉浦 亮介 東京都千代田区大手町一丁目5番1号 日 本電信電話株式会社内

最終頁に続く

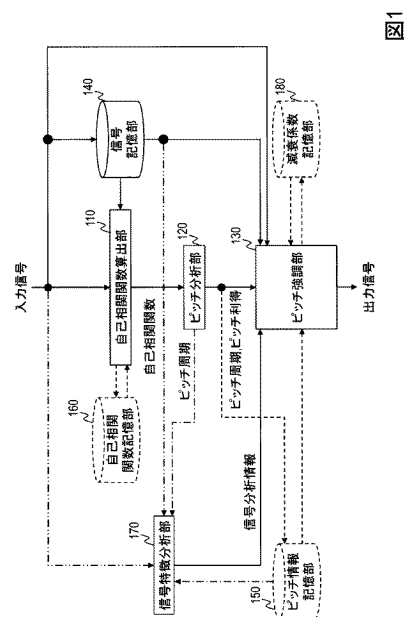
(54) 【発明の名称】 ピッチ強調装置、その方法、およびプログラム

(57) 【要約】

【課題】 子音の時間区間であっても違和感が少ないピッチ強調処理であり、子音の時間区間とそれ以外の時間区間とが頻繁に切り替わる場合であっても不連続に基づく受聴時の違和感の少ないピッチ強調処理を実現する。

【解決手段】 ピッチ強調装置は、ピッチ強調処理として、信号のスペクトル包絡が平坦であると判定された時間区間については、当該時間区間の各時刻について、当該時間区間のピッチ周期に対応するサンプル数 T だけ、当該時刻よりも過去の時刻の信号と、当該時間区間のピッチ利得 G と、所定の定数 B と、0より大きく1より小さい値 α と、を乗算した信号と、当該時刻の信号と、を加算した信号を含む信号を出力信号として得る。

【選択図】 図1



【特許請求の範囲】

【請求項 1】

入力された音信号に由来する信号に対して時間区間毎にピッチ強調処理を施して出力信号を得るピッチ強調装置であって、

前記ピッチ強調処理として、

前記信号のスペクトル包絡が平坦であると判定された時間区間については、

当該時間区間の各時刻について、

当該時間区間のピッチ周期に対応するサンプル数 T_i だけ、当該時刻よりも過去の時刻の前記信号と、当該時間区間のピッチ利得 G_i と、所定の定数 B_i と、0より大きく1より小さい値と、を乗算した信号と、

10

当該時刻の前記信号と、を加算した信号を含む信号を出力信号として得、

前記信号のスペクトル包絡が平坦でないと判定された時間区間については、

当該時間区間の各時刻について、

当該時間区間のピッチ周期に対応するサンプル数 T_i だけ、当該時刻よりも過去の時刻の前記信号と、当該時間区間のピッチ利得 G_i と、所定の定数 B_i と、を乗算した信号と、

当該時刻の前記信号と、を加算した信号を含む信号を出力信号として得る処理を行うピッチ強調部を含む、

ピッチ強調装置。

【請求項 2】

入力された音信号に由来する信号に対して時間区間毎にピッチ強調処理を施して出力信号を得るピッチ強調装置であって、

20

前記ピッチ強調処理として、

前記各時間区間の各時刻 n について、

当該時間区間のピッチ周期に対応するサンプル数 T_i だけ、当該時刻 n よりも過去の時刻の前記信号と、当該時間区間のピッチ利得 G_i と、当該時間区間のスペクトル包絡が平坦であればあるほど小さい値と、を乗算した信号と、

当該時刻 n の前記信号と、を加算した信号を含む信号を出力信号として得る処理を行うピッチ強調部を含む、

ピッチ強調装置。

【請求項 3】

30

入力された音信号に由来する信号に対して時間区間毎にピッチ強調処理を施して出力信号を得るピッチ強調方法であって、

前記ピッチ強調処理として、

前記信号のスペクトル包絡が平坦であると判定された時間区間については、

当該時間区間の各時刻について、

当該時間区間のピッチ周期に対応するサンプル数 T_i だけ、当該時刻よりも過去の時刻の前記信号と、当該時間区間のピッチ利得 G_i と、所定の定数 B_i と、0より大きく1より小さい値と、を乗算した信号と、

当該時刻の前記信号と、を加算した信号を含む信号を出力信号として得、

前記信号のスペクトル包絡が平坦でないと判定された時間区間については、

40

当該時間区間の各時刻について、

当該時間区間のピッチ周期に対応するサンプル数 T_i だけ、当該時刻よりも過去の時刻の前記信号と、当該時間区間のピッチ利得 G_i と、所定の定数 B_i と、を乗算した信号と、

当該時刻の前記信号と、を加算した信号を含む信号を出力信号として得る処理を行うピッチ強調ステップを含む、

ピッチ強調方法。

【請求項 4】

入力された音信号に由来する信号に対して時間区間毎にピッチ強調処理を施して出力信号を得るピッチ強調方法であって、

前記ピッチ強調処理として、

50

前記各時間区間の各時刻 n について、

当該時間区間のピッチ周期に対応するサンプル数 T_0 だけ、当該時刻 n よりも過去の時刻の前記信号と、当該時間区間のピッチ利得 G と、当該時間区間のスペクトル包絡が平坦であればあるほど小さい値と、を乗算した信号と、

当該時刻 n の前記信号と、を加算した信号を含む信号を出力信号として得る処理を行うピッチ強調ステップを含む、

ピッチ強調方法。

【請求項 5】

請求項 1 または請求項 2 のピッチ強調装置としてコンピュータを機能させるためのプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

この発明は、音信号の符号化技術などの信号処理技術において、音信号に由来するサンプル列に対して、そのピッチ成分を分析し、強調する技術に関連する。

【背景技術】

【0002】

一般的に、時系列信号などのサンプル列を非可逆に圧縮符号化した場合、復号時に得られるサンプル列は元のサンプル列とは違った、歪のあるサンプル列となる。特に音信号の符号化においては、この歪が自然音にはないようなパターンを含むことが多く、復号した音信号を受聴した際に不自然に感じられることがある。そこで、自然音の多くがある一定区間で観測した際に音に応じた周期成分、つまりピッチを含むことに着目し、復号により得た音信号の各サンプルに対して、ピッチ周期分だけ過去のサンプルを加算することにより、ピッチ成分を強調する処理を行い、より違和感の少ない音に変換する技術が広く用いられている（例えば非特許文献 1）。

【0003】

また、例えば特許文献 1 に記載されているように、復号により得た音信号が「音声」であるか「非音声」であるかの情報に基づき、「音声」である場合にはピッチ成分を強調する処理を行い、「非音声」である場合にはピッチ成分を強調する処理を行わない技術もある。

【先行技術文献】

【非特許文献】

【0004】

【非特許文献 1】ITU T Recommendation G.723.1 (05/2006) pp.16 18, 2006

【特許文献】

【0005】

【特許文献 1】特開平 10 - 143195 号公報

【発明の概要】

【発明が解決しようとする課題】

【0006】

しかしながら、非特許文献 1 に記載された技術には、明確なピッチ構造をもたない子音部についてもピッチ成分を強調する処理を行ってしまうことにより、子音部を受聴した際に不自然に感じられるという課題がある。一方、特許文献 1 に記載された技術では、子音部に信号としてはピッチ成分が存在している場合であってもピッチ成分を強調する処理を全く行わないことから、子音部を受聴した際に不自然に感じられるという課題がある。また、特許文献 1 に記載された技術には、母音の時間区間と子音の時間区間とでピッチ強調処理の有無が切り替わることによって音信号に不連続が頻繁に生じてしまい、受聴時の違和感が増してしまう、という課題もある。

【0007】

本発明は、これらの課題を解決するためのものであり、子音の時間区間であっても違和

10

20

30

40

50

感が少ないピッチ強調処理であり、子音の時間区間とそれ以外の時間区間とが頻繁に切り替わる場合であっても不連続に基づく受聴時の違和感の少ないピッチ強調処理を実現することを目的とする。なお、子音は、摩擦音、破裂音、半母音、鼻音、および破擦音を含む(参考文献1、参考文献2参照)。

(参考文献1)古井貞熙著、「音響・音声工学」、近代科学社、1992年、p.99

(参考文献2)斎藤収三、中田和男、「音声情報処理の基礎」、オーム社、1981年、p.38 3

9

【課題を解決するための手段】

【0008】

上記の課題を解決するために、本発明の一態様によれば、ピッチ強調装置は、入力された音信号に由来する信号に対して時間区間毎にピッチ強調処理を施して出力信号を得る。ピッチ強調装置は、ピッチ強調処理として、信号のスペクトル包絡が平坦であると判定された時間区間については、当該時間区間の各時刻について、当該時間区間のピッチ周期に対応するサンプル数 T_i だけ、当該時刻よりも過去の時刻の信号と、当該時間区間のピッチ利得 G_i と、所定の定数 B_i と、0より大きく1より小さい値と、を乗算した信号と、当該時刻の信号と、を加算した信号を含む信号を出力信号として得、信号のスペクトル包絡が平坦でない判定された時間区間については、当該時間区間の各時刻について、当該時間区間のピッチ周期に対応するサンプル数 T_i だけ、当該時刻よりも過去の時刻の信号と、当該時間区間のピッチ利得 G_i と、所定の定数 B_i と、を乗算した信号と、当該時刻の信号と、を加算した信号を含む信号を出力信号として得る処理を行うピッチ強調部を含む。

10

20

【0009】

上記の課題を解決するために、本発明の他の態様によれば、ピッチ強調装置は、入力された音信号に由来する信号に対して時間区間毎にピッチ強調処理を施して出力信号を得る。ピッチ強調装置は、ピッチ強調処理として、各時間区間の各時刻 n について、当該時間区間のピッチ周期に対応するサンプル数 T_i だけ、当該時刻 n よりも過去の時刻の信号と、当該時間区間のピッチ利得 G_i と、当該時間区間のスペクトル包絡が平坦であればあるほど小さい値と、を乗算した信号と、当該時刻 n の信号と、を加算した信号を含む信号を出力信号として得る処理を行うピッチ強調部を含む。

【発明の効果】

【0010】

本発明によれば、復号処理により得られた音声信号に対してピッチ強調処理を施す場合に、子音の時間区間であっても違和感が少なく、子音の時間区間とそれ以外の時間区間とが頻繁に切り替わる場合であっても不連続に基づく受聴時の違和感の少ないピッチ強調処理を実現することができるという効果を奏する。

30

【図面の簡単な説明】

【0011】

【図1】第一実施形態、第二実施形態、第三実施形態、およびそれらの変形例に係るピッチ強調装置の機能ブロック図。

【図2】第一実施形態、第二実施形態、第三実施形態、およびそれらの変形例に係るピッチ強調装置の処理フローの例を示す図。

40

【図3】その他の変形例に係るピッチ強調装置の機能ブロック図。

【図4】その他の変形例に係るピッチ強調装置の処理フローの例を示す図。

【発明を実施するための形態】

【0012】

以下、本発明の実施形態について、説明する。なお、以下の説明に用いる図面では、同じ機能を持つ構成部や同じ処理を行うステップには同一の符号を記し、重複説明を省略する。以下の説明において、ベクトルや行列の各要素単位で行われる処理は、特に断りがない限り、そのベクトルやその行列の全ての要素に対して適用されるものとする。

【0013】

<第一実施形態>

50

図 1 は第一実施形態に係る音声ピッチ強調装置の機能ブロック図を、図 2 はその処理フローを示す。

【 0 0 1 4 】

図 1 を参照して、第一実施形態の音声ピッチ強調装置の処理手続きを説明する。第一実施形態の音声ピッチ強調装置は、入力された信号を分析してピッチ周期とピッチ利得を得て、そのピッチ周期とピッチ利得に基づきピッチを強調するものである。本実施形態では、時間区間ごとの入力された音信号に対してピッチ周期に対応するピッチ成分にピッチ利得を乗算したものをを用いてピッチ強調処理を施す際に、スペクトル包絡が平坦である時間区間のピッチ成分の強調の度合いを、スペクトル包絡が平坦でない時間区間のピッチ成分の強調の度合いより小さくする。または、時間区間のピッチ成分の強調の度合いを、スペクトル包絡が平坦であればあるほど小さくする。より具体的には、スペクトル包絡が平坦である時間区間については、ピッチ利得の代わりに、ピッチ利得に1より小さい値を乗算したものをを用いる。子音のスペクトルは、母音に比べ、スペクトル包絡が平坦になる性質がある。本実施形態では、上述の課題を解決するために、この性質を利用し、強調の度合いを変更する。

10

【 0 0 1 5 】

第一実施形態の音声ピッチ強調装置は、信号特徴分析部 1 7 0 と自己相関関数算出部 1 1 0 とピッチ分析部 1 2 0 とピッチ強調部 1 3 0 と信号記憶部 1 4 0 とを備えるものであり、更にピッチ情報記憶部 1 5 0 と自己相関関数記憶部 1 6 0 と減衰係数記憶部 1 8 0 とを備えてもよい。

20

【 0 0 1 6 】

音声ピッチ強調装置は、例えば、中央演算処理装置 (CPU: Central Processing Unit)、主記憶装置 (RAM: Random Access Memory) などをも有する公知又は専用のコンピュータに特別なプログラムが読み込まれて構成された特別な装置である。音声ピッチ強調装置は、例えば、中央演算処理装置の制御のもとで各処理を実行する。音声ピッチ強調装置に入力されたデータや各処理で得られたデータは、例えば、主記憶装置に格納され、主記憶装置に格納されたデータは必要に応じて中央演算処理装置へ読み出されて他の処理に利用される。音声ピッチ強調装置の各処理部は、少なくとも一部が集積回路等のハードウェアによって構成されていてもよい。音声ピッチ強調装置が備える各記憶部は、例えば、RAM (Random Access Memory) などの主記憶装置、またはリレーショナルデータベースやキーバリューストアなどのミドルウェアにより構成することができる。ただし、各記憶部は、必ずしも音声ピッチ強調装置がその内部に備える必要はなく、ハードディスクや光ディスクもしくはフラッシュメモリ (Flash Memory) のような半導体メモリ素子により構成される補助記憶装置により構成し、音声ピッチ強調装置の外部に備える構成としてもよい。

30

【 0 0 1 7 】

第一実施形態の音声ピッチ強調装置が行う主な処理は自己相関関数算出処理 (S 1 1 0) とピッチ分析処理 (S 1 2 0) と信号特徴分析処理 (S 1 7 0) とピッチ強調処理 (S 1 3 0) であり (図 2 参照)、これらの処理は音声ピッチ強調装置が備える複数のハードウェア資源が連携して行うものである。以下では、自己相関関数算出処理 (S 1 1 0) とピッチ分析処理 (S 1 2 0) と信号特徴分析処理 (S 1 7 0) とピッチ強調処理 (S 1 3 0) のそれぞれについて、関連する処理と共に説明する。

40

【 0 0 1 8 】

[自己相関関数算出処理 (S 1 1 0)]

まず、音声ピッチ強調装置が行う自己相関関数算出処理とこれに関連する処理について説明する。

【 0 0 1 9 】

自己相関関数算出部 1 1 0 には、時間領域の音信号(入力信号)が入力される。この音信号は、例えば音声信号などの音響信号を符号化装置で圧縮符号化して符号を得て、その符号化装置に対応する復号装置で符号を復号して得た信号である。自己相関関数算出部 1 1 0 には、所定の時間長のフレーム (時間区間) 単位で、音声ピッチ強調装置に入力された

50

現在のフレームの時間領域の音信号のサンプル列が入力される。1フレームのサンプル列の長さを示す正の整数をNとすると、自己相関関数算出部110には、現在のフレームの時間領域の音信号のサンプル列を構成するN個の時間領域の音信号サンプルが入力される。自己相関関数算出部110は、入力されたN個の時間領域の音信号サンプルを含む最新のL個(Lは正の整数)の音信号サンプルによるサンプル列における時間差0の自己相関関数 R_0 及び複数個(M個、Mは正の整数)の所定の時間差(1), ..., (M)それぞれに対する自己相関関数 $R_{(1)}$, ..., $R_{(M)}$ を算出する。すなわち、自己相関関数算出部110は、現在のフレームの時間領域の音信号サンプルを含む最新の音信号サンプルによるサンプル列における自己相関関数を算出する。

【0020】

なお、以降では、現在のフレームの処理において自己相関関数算出部110が算出した自己相関関数、すなわち、現在のフレームの時間領域の音信号サンプルを含む最新の音信号サンプルによるサンプル列における自己相関関数、のことを「現在のフレームの自己相関関数」とも呼ぶ。同様に、過去のあるフレームをフレームFとしたとき、フレームFの処理において自己相関関数算出部110が算出した自己相関関数、すなわち、フレームFの時間領域の音信号サンプルを含むフレームFの時点での最新の音信号サンプルによるサンプル列における自己相関関数、のことを「フレームFの自己相関関数」とも呼ぶ。また、「自己相関関数」は単に「自己相関」と呼ぶこともある。LがNより大きい値である場合には、自己相関関数の算出に最新のL個の音信号サンプルを用いるために、音声ピッチ強調装置内には信号記憶部140を備え、1つ前のフレームまでに入力された最新の少なくともL-N個の音信号サンプルを記憶できるようにしておく。そして、自己相関関数算出部110は、現在のフレームのN個の時間領域の音信号サンプルが入力された際には、信号記憶部140に記憶された最新のL-N個の音信号サンプルを $X_0, X_1, \dots, X_{L-N-1}$ として読み出し、入力されたN個の時間領域の音信号サンプルを $X_{L-N}, X_{L-N+1}, \dots, X_{L-1}$ とすることにより、最新のL個の音信号サンプル X_0, X_1, \dots, X_{L-1} を得る。

【0021】

そして、自己相関関数算出部110は、最新のL個の音信号サンプル X_0, X_1, \dots, X_{L-1} を用いて、時間差0の自己相関関数 R_0 、及び複数個の所定の時間差(1), ..., (M)それぞれに対する自己相関関数 $R_{(1)}$, ..., $R_{(M)}$ を算出する。(1), ..., (M)や0などの時間差を

とすると、自己相関関数算出部110は、自己相関関数 R を例えば以下の式(1)で算出する。

【0022】

【数1】

$$R_{\tau} = \sum_{l=\tau}^{L-1} X_l X_{l-\tau} \quad \dots(1)$$

【0023】

自己相関関数算出部110は算出した自己相関関数 $R_0, R_{(1)}, \dots, R_{(M)}$ をピッチ分析部120に出力する。

【0024】

なお、この時間差(1), ..., (M)は後述するピッチ分析部120が求める現在のフレームのピッチ周期 T_0 の候補である。例えば、サンプリング周波数32kHzの音声信号を主とする音信号の場合には、音声のピッチ周期の候補として好適な75から320までの整数値を(1), ..., (M)とするなどの実装が考えられる。なお、式(1)の R に代えて、式(1)の R を R_0 で除算した正規化自己相関関数 R/R_0 を求めてもよい。ただし、Lを8192などのピッチ周期 T_0 の候補である75から320に対して十分に大きな値とした場合などには、自己相関関数 R に代えて正規化自己相関関数 R/R_0 を求めるよりも、以下で説明する演算量を抑えた方法で自己相関関数 R を算出するほうがよい。

【0025】

自己相関関数 R は、式(1)そのもので算出してもよいが、式(1)で求まるのと同じ値を

10

20

30

40

50

別の算出方法で算出してもよい。例えば、音声ピッチ強調装置内に自己相関関数記憶部 160 を備えて 1 つ前のフレーム（直前のフレーム）の自己相関関数を算出する処理で得られた自己相関関数（直前のフレーム自己相関関数） $R_{(1)}$ 、 $R_{(M)}$ を記憶しておき、自己相関関数算出部 110 は、自己相関関数記憶部 160 から読み出した直前のフレームの処理で得られた自己相関関数（直前のフレーム自己相関関数） $R_{(1)}$ 、 $R_{(M)}$ それぞれに、新たに入力された現在のフレームの音信号サンプルの寄与分の加算と、最も過去のフレームの寄与分の減算と、を行うことにより現在のフレームの自己相関関数 $R_{(1)}$ 、 $R_{(M)}$ を算出するようにしてもよい。これにより、式(1)そのもので算出するよりも自己相関関数の算出に要する演算量を抑えることが可能である。この場合、(1)、(M)のそれぞれを $R_{(1)}$ とすると、自己相関関数算出部 110 は、現在のフレームの自己相関関数 $R_{(1)}$ を、直前のフレームの処理で得られた自己相関関数 $R_{(1)}$ （直前のフレームの自己相関関数 $R_{(1)}$ ）に対して、以下の式(2)で得られる差分 ΔR_{τ}^{+} を加算し、式(3)で得られる差分 ΔR_{τ}^{-} を減算することにより得る。

10

【0026】

【数2】

$$\Delta R_{\tau}^{+} = \sum_{l=L-N}^{L-1} X_l X_{l-\tau} \quad \dots(2)$$

$$\Delta R_{\tau}^{-} = \sum_{l=\tau}^{N-1+\tau} X_l X_{l-\tau} \quad \dots(3)$$

【0027】

また、入力された音信号の最新の L 個の音信号サンプルそのものではなく、当該 L 個の音信号サンプルに対してダウンサンプリングやサンプルの間引きなどを行うことによりサンプル数を減らした信号を用いて、上記と同様の処理により自己相関関数を算出することで演算量を節約してもよい。この場合、M 個の時間差 (1)、(M) は、例えばサンプル数を半分にした際には半分のサンプル数で表現する。例えば、上述したサンプリング周波数 32kHz の 8192 個の音信号サンプルをサンプリング周波数 16kHz の 4096 個のサンプルにダウンサンプリングした場合には、ピッチ周期 T の候補である (1)、(M) は、75 から 320 の約半分である 37 から 160 とすればよい。

30

【0028】

なお、信号記憶部 140 に記憶された音信号サンプルは後述する信号特徴分析処理にも用いられる。具体的には、後述する信号特徴分析処理では信号記憶部 140 に記憶された J N 個（J は正の整数）の音信号サンプルが用いられる。すなわち、L と J のうちの大きいほうの値を K とすると（ $K = \max(L, J)$ とすると）、1 つ前のフレームまでに入力された最新の少なくとも K - N 個の音信号サンプルを信号記憶部 140 に記憶しておく必要がある。そこで、信号記憶部 140 は、音声ピッチ強調装置が現在のフレームについての後述するピッチ強調部 130 の処理までを終えた後に、その時点で最新の K - N 個の音信号サンプルを記憶しておくように記憶内容を更新する。具体的には、例えば、 $K > 2N$ の場合、信号記憶部 140 は、記憶されている K - N 個の音信号サンプルのうち一番古い N 個の音信号サンプル $XR_0, XR_1, \dots, XR_{N-1}$ を削除し、 $XR_N, XR_{N+1}, \dots, XR_{K-N-1}$ を $XR_0, XR_1, \dots, XR_{K-2N-1}$ とし、入力された現在のフレームの N 個の時間領域の音信号サンプルを $XR_{K-2N}, XR_{L-2N+1}, \dots, XR_{K-N-1}$ として新たに記憶する。また、 $K = 2N$ の場合、信号記憶部 140 は、記憶されている K - N 個の音信号サンプル $XR_0, XR_1, \dots, XR_{K-N-1}$ を削除し、入力された現在のフレームの N 個の時間領域の音信号サンプルのうち最新の K - N 個の音信号サンプルを $XR_0, XR_1, \dots, XR_{K-N-1}$ として新たに記憶する。なお、 $K = N$ である場合には、音声ピッチ強調装置内には信号記憶部 140 を備える必要はない。

40

【0029】

また、自己相関関数記憶部 160 は、自己相関関数算出部 110 が現在のフレームについての自己相関関数の算出を終えた後に、算出した現在のフレームの自己相関関数 $R_{(1)}$

50

, $R_{(M)}$ を記憶しておくように記憶内容を更新する。具体的には、自己相関関数記憶部 160は、記憶されている $R_{(1)}$, $R_{(M)}$ を削除し、算出した現在のフレームの自己相関関数 $R_{(1)}$, $R_{(M)}$ を新たに記憶する。

【0030】

なお、上述の説明では、最新のL個の音信号サンプルが現在のフレームのN個の音信号サンプルを含む(つまり $L = N$)ことを前提としているが、必ずしも $L = N$ である必要はなく、 $L < N$ であってもよい。この場合、自己相関関数算出部 110は、現在のフレームのN個に含まれる連続したL個の音信号サンプル $X_i, X_{i+1}, \dots, X_{i+L-1}$ を用いて、時間差0の自己相関関数 R_0 、及び複数個の所定の時間差(1), \dots , (M)それぞれに対する自己相関関数 $R_{(1)}$, $R_{(M)}$ を算出すればよい。

10

【0031】

[ピッチ分析処理(S120)]

次に、音声ピッチ強調装置が行うピッチ分析処理について説明する。

【0032】

ピッチ分析部 120には、自己相関関数算出部 110が出力した現在のフレームの自己相関関数 $R_0, R_{(1)}, \dots, R_{(M)}$ が入力される。

【0033】

ピッチ分析部 120は、所定の時間差に対する現在のフレームの自己相関関数 $R_{(1)}, \dots, R_{(M)}$ の中での最大値を求め、自己相関関数の最大値と時間差0の自己相関関数 R_0 の比を現在のフレームのピッチ利得 I_1 として得て、また、自己相関関数が最大値となる時間差を現在のフレームのピッチ周期 T_1 として得て、それぞれをピッチ強調部 130へ出力する。

20

【0034】

[信号特徴分析処理(S170)]

次に、音声ピッチ強調装置が行う信号特徴分析処理について説明する。

【0035】

信号特徴分析部 170には、時間領域の音信号に由来する情報が入力される。この音信号は、自己相関関数算出部 110に入力される音信号と同じ信号である。

【0036】

例えば、信号特徴分析部 170には、所定の時間長のフレーム(時間区間)単位で、音声ピッチ強調装置に入力された現在のフレームの時間領域の音信号のサンプル列が入力される。すなわち、信号特徴分析部 170には、現在のフレームの時間領域の音信号のサンプル列を構成するN個の時間領域の音信号サンプルが入力される。この場合には、信号特徴分析部 170は、入力されたN個の時間領域の音信号サンプルを含む最新のJ個(Jは正の整数)の音信号サンプルによるサンプル列を用いて、現在のフレームのスペクトル包絡が平坦であるか否かを表す情報、または、現在のフレームのスペクトル包絡の平坦度合いの指標値、を得て、信号分析情報 I_2 としてピッチ強調部 130へ出力する。すなわち、この場合は、「時間領域の音信号に由来する情報」は現在のフレームの時間領域の音信号のサンプル列(図1中、二点鎖線で示す)である。

30

【0037】

なお、前述の通り、子音のスペクトルは、母音に比べ、スペクトル包絡が平坦になる性質がある。そこで、「スペクトル包絡の平坦度合いの指標値」を「子音らしさの指標値」ともいい、「スペクトル包絡が平坦であるか否かを表す情報」を「子音であるか否かを表す情報」ともいう。

40

【0038】

信号特徴分析部 170は例えば下記の例1-1から例1-7の信号特徴分析処理により信号分析情報 I_2 を得る。

【0039】

(信号特徴分析処理の例1-1:スペクトル包絡の平坦度合いの指標値を信号分析情報とする例その1)

50

この例では、信号特徴分析部 170 は、まず、入力されたN個の時間領域の音信号サンプルを含む最新のJ個の音信号サンプルによるサンプル列からT次のLSPパラメータ [1], [2], ..., [T]を得る (Step 111)。信号特徴分析部 170 は、次に、Step 111で得たT次のLSPパラメータ [1], [2], ..., [T]を用いて、下記の指標Qを現在のフレームのスペクトル包絡の平坦度合いの指標値 (便宜上、「子音らしさの第1 - 1の指標値」ともいう)として得る (Step 112)。

【0040】

【数3】

$$Q = \frac{1}{\frac{1}{(T-1)} \sum_i^{T-1} (\bar{\theta} - \theta[i+1] + \theta[i])^2} \quad \dots(11)$$

$$\text{ただし、} \bar{\theta} = \frac{1}{(T-1)} \sum_i^{T-1} (\theta[i+1] - \theta[i])$$

【0041】

(信号特徴分析処理の例1 - 2 : スペクトル包絡の平坦度合いの指標値を信号分析情報とする例その2)

この例では、信号特徴分析部 170 は、まず、入力されたN個の時間領域の音信号サンプルを含む最新のJ個の音信号サンプルによるサンプル列からT次のLSPパラメータ [1], [2], ..., [T]を得る (Step 121)。信号特徴分析部 170 は、次に、Step 121で得たT次のLSPパラメータ [1], [2], ..., [T]を用いて、隣接するLSPパラメータの間隔の最小値、すなわち、下記の指標Q'、を現在のフレームのスペクトル包絡の平坦度合いの指標値 (便宜上、「子音らしさの第1 - 2の指標値」ともいう)として得る (Step 122)

20

【0042】

【数4】

$$Q' = \min_{i \in \{1, \dots, T-1\}} (\theta[i+1] - \theta[i]) \quad \dots(12)$$

【0043】

(信号特徴分析処理の例1 - 3 : スペクトル包絡の平坦度合いの指標値を信号分析情報とする例その3)

この例では、信号特徴分析部 170 は、まず、入力されたN個の時間領域の音信号サンプルを含む最新のJ個の音信号サンプルによるサンプル列からT次のLSPパラメータ [1], [2], ..., [T]を得る (Step 131)。信号特徴分析部 170 は、次に、Step 131で得たT次のLSPパラメータ [1], [2], ..., [T]を用いて、隣接するLSPパラメータの間隔の値と最低次のLSPパラメータの値のうち、すなわち、下記の指標Q''、を現在のフレームのスペクトル包絡の平坦度合いの指標値 (便宜上、「子音らしさの第1 - 3の指標値」ともいう)として得る (Step 132)。

30

【0044】

【数5】

$$Q'' = \min_{i \in \{1, \dots, T-1\}} (\min(\theta[i+1] - \theta[i], \theta[1])) \quad \dots(13)$$

40

【0045】

(信号特徴分析処理の例1 - 4 : スペクトル包絡の平坦度合いの指標値を信号分析情報とする例その4)

この例では、信号特徴分析部 170 は、まず、入力されたN個の時間領域の音信号サンプルを含む最新のJ個の音信号サンプルによるサンプル列からp次のPARCOR係数k[1], k[2], ..., k[p]を得る (Step 141)。信号特徴分析部 170 は、次に、Step 141で得たp次のPARCOR係数k[1], k[2], ..., k[p]を用いて、下記の指標Q'''を現在のフレームのスペクトル

50

包絡の平坦度合いの指標値（便宜上、「子音らしさの第 1 - 4 の指標値」ともいう）として得る（Step 1 4 2）。

【 0 0 4 6 】

【 数 6 】

$$Q'' = \prod_i^P (1 - k[i]^2) \quad \dots(14)$$

【 0 0 4 7 】

（信号特徴分析処理の例 1 - 5：複数の指標値を組み合わせた指標値を信号分析情報とする例）

この例では、信号特徴分析部 1 7 0 は、例 1 - 1 ~ 例 1 - 4 の方法により、子音らしさの第 1 - 1 ~ 第 1 - 4 の指標値を得る（Step 1 5 1）。信号特徴分析部 1 7 0 は、さらに、Step 1 5 1 で得た子音らしさの第 1 - 1 ~ 第 1 - 4 の指標値の重み付け加算により、第 1 - 1 の指標値が大きな値になるほど大きな値になり、かつ、第 1 - 2 の指標値が大きな値になるほど大きな値になり、かつ、第 1 - 3 の指標値が大きな値になるほど大きな値になり、かつ、第 1 - 4 の指標値が大きな値になるほど大きな値になる値を、現在のフレームの子音らしさの指標値（便宜上、「第 1 - 5 の指標値」ともいう）として得て、得た第 1 - 5 の指標値を信号分析情報₁として出力する（Step 1 5 2）。

【 0 0 4 8 】

前述の通り、子音らしさの第 1 - 1 ~ 第 1 - 4 の指標値は、それぞれ子音らしさを表す指標である。この例では 4 つの指標値を組み合わせることでより柔軟に子音らしさの指標値を設定することができる。

【 0 0 4 9 】

なお、信号特徴分析部 1 7 0 は、子音らしさの第 1 - 1 ~ 第 1 - 4 の指標値のうちの少なくとも 2 個を得て（Step 1 5 1'）、Step 1 5 1' で得た少なくとも 2 個の子音らしさの指標値の重み付け加算により、Step 1 5 1' で得た指標値それぞれが大きな値になるほど大きな値になる値を、現在のフレームの子音らしさの第 1 - 5 の指標値として得て、得た第 1 - 5 の指標値を信号分析情報₁として出力してもよい（Step 1 5 2'）。

【 0 0 5 0 】

信号特徴分析処理の例 1 - 1 ~ 例 1 - 5 ではスペクトル包絡の平坦度合いの指標値（子音らしさの指標値）を信号分析情報とする例を説明してきた。ここからは、スペクトル包絡が平坦であるか否かを表す情報（子音であるか否かを表す情報）を信号分析情報とする例を説明する。

【 0 0 5 1 】

（信号特徴分析処理の例 1 - 6：スペクトル包絡が平坦であるか否かを表す情報を信号分析情報とする例その 1）

この例では、信号特徴分析部 1 7 0 は、まず、例 1 - 1 ~ 例 1 - 5 の何れかと同じ方法により、現在のフレームの子音らしさの第 1 - 1 ~ 第 1 - 5 の指標値の何れかを得る（Step 1 6 1）。信号特徴分析部 1 7 0 は、次に、Step 1 6 1 で得た指標値が予め定めた閾値以上または閾値を超える場合には、現在のフレームが子音であることを表す情報（「第 1 - 1 の指標値」~「第 1 - 5 の指標値」に対応する「現在のフレームが子音であるか否かを表す情報」をそれぞれ、便宜上、「第 1 - 1 の情報」~「第 1 - 5 の情報」ともいう）を信号分析情報₁として出力し、そうでない場合には、現在のフレームが子音でないことを表す第 2 - 1 ~ 第 2 - 5 の情報の何れかを信号分析情報₁として出力する（Step 1 6 2）。

【 0 0 5 2 】

（信号特徴分析処理の例 1 - 7：スペクトル包絡が平坦であるか否かを表す情報を信号分析情報とする例その 2）

この例では、信号特徴分析部 1 7 0 は、まず、例 1 - 1 ~ 例 1 - 4 と同じ方法により、現在のフレームの子音らしさの第 1 - 1 ~ 第 1 - 4 の指標値を得る（Step 1 7 1）。次に

10

20

30

40

50

、信号特徴分析部 170 は、Step 171 で得た 4 つの子音らしさの第 1 - 1 ~ 第 1 - 4 の指標値のそれぞれと、予め定めた閾値との大小関係に基づき、各子音らしさの第 1 - 1 ~ 第 1 - 4 の指標値に対して、現在のフレームが子音であることを表す情報、または、現在のフレームが子音でないことを表す情報を得る (Step 172)。なお、閾値は 4 つの第 1 - 1 ~ 第 1 - 4 の指標値毎に設定するものとし、第 1 - 1 ~ 第 1 - 4 の指標値に対応する現在のフレームが子音であるか否かを表す情報をそれぞれ第 1 - 1 ~ 第 1 - 4 の情報ともいう。例えば、第 1 - 1 の指標値が予め定めた閾値以上または閾値を超える場合には、現在のフレームが子音であることを表す第 1 - 1 の情報を得て、そうでない場合には、現在のフレームが子音でないことを表す第 1 - 1 の情報を得る。同様に、第 1 - 2 ~ 第 1 - 4 の指標値と予め定めた閾値との大小関係に基づき第 1 - 2 ~ 第 1 - 4 の情報を得る。

10

【0053】

信号特徴分析部 170 は、4 つの第 1 - 1 ~ 第 1 - 4 の情報の論理演算に基づき、現在のフレームが子音であることを表す情報 (便宜上、「第 1 - 6 の情報」ともいう)、または、現在のフレームが子音でないことを表す第 1 - 6 の情報を得る (Step 173)。

【0054】

(論理演算の例 1)

例えば、第 1 - 1 ~ 第 1 - 4 の情報全てが子音であることを表す場合には、現在のフレームが子音であることを表す第 1 - 6 の情報を信号分析情報 I_6 として出力し、そうでない場合には、現在のフレームが子音でないことを表す第 1 - 6 の情報を信号分析情報 I_6 として出力する。

20

【0055】

(論理演算の例 2)

また、例えば、第 1 - 1 ~ 第 1 - 4 の情報の何れかが子音であることを表す場合には、現在のフレームが子音であることを表す第 1 - 6 の情報を信号分析情報 I_6 として出力し、そうでない場合には、現在のフレームが子音でないことを表す第 1 - 6 の情報を信号分析情報 I_6 として出力する。

【0056】

(論理演算の例 3)

また、例えば、第 1 - 1 ~ 第 1 - 2 の情報の何れかが子音であることを表し、かつ、第 1 - 3 ~ 第 1 - 4 の情報の何れかが子音であることを表す場合 (論理和と論理積の組合せを用いる場合) には、現在のフレームが子音であることを表す第 1 - 6 の情報を信号分析情報 I_6 として出力し、そうでない場合には、現在のフレームが子音でないことを表す第 1 - 6 の情報を信号分析情報 I_6 として出力する。

30

【0057】

なお、第 1 - 1 ~ 第 1 - 4 の情報の論理演算は上述の論理演算の例 1 ~ 3 に限定されるものではなく、復号した音信号がより自然に感じられるように適宜設定すればよい。

【0058】

また、信号特徴分析部 170 は、子音らしさの第 1 - 1 ~ 第 1 - 4 の指標値のうち少なくとも 2 個を得て (Step 171')、Step 171' で得た少なくとも 2 個の子音らしさの指標値のそれぞれと、予め定めた閾値との大小関係に基づき、各子音らしさの指標値に対して、現在のフレームが子音であることを表す情報、または、現在のフレームが子音でないことを表す少なくとも 2 個の情報を得て (Step 172')、Step 172' で得た少なくとも 2 個の情報の論理演算に基づき、現在のフレームが子音であることを表す第 1 - 6 の情報、または、現在のフレームが子音でないことを表す第 1 - 6 の情報を得てもよい (Step 173')。

40

【0059】

このような処理により信号特徴分析部 170 は、子音らしさの指標値または子音であるか否かを表す情報を信号分析情報 I_6 として出力する。

【0060】

[ピッチ強調処理 (S130)]

50

次に、音声ピッチ強調装置が行うピッチ強調処理について説明する。

【0061】

ピッチ強調部130は、ピッチ分析部120が出力したピッチ周期とピッチ利得、信号特徴分析部170が出力した信号分析情報、及び音声ピッチ強調装置に入力された現在のフレームの時間領域の音信号(入力信号)を受け取り、現在のフレームの音信号サンプル列に対し、現在のフレームのピッチ周期 T_0 に対応するピッチ成分を、ピッチ利得 β に基づく強調の度合いが子音のフレーム(スペクトル包絡が平坦であるフレーム)のほうが子音以外のフレーム(スペクトル包絡が平坦でないフレーム)よりも小さくなるように、強調して得た出力信号のサンプル列を出力する。

【0062】

以下、具体例を説明する。

【0063】

ピッチ強調部130は、入力された現在のフレームのピッチ利得 β と、入力された現在のフレームのピッチ周期 T_0 と、入力された現在のフレームの信号分析情報 I_0 とを用い、現在のフレームの音信号のサンプル列に対するピッチ強調処理を行う。具体的には、ピッチ強調部130は、入力された現在のフレームの音信号のサンプル列を構成する各サンプル X_n ($L-N \leq n \leq L-1$)に対して、以下の式(21)により出力信号 X_n^{new} を得ることにより、 N 個のサンプル $X_{L-N}^{new}, \dots, X_{L-1}^{new}$ による現在のフレームの出力信号のサンプル列を得る。

【0064】

【数7】

$$X_n^{new} = \frac{1}{A} \left[X_n + B_0 \sigma_0 \gamma_0 X_{n-T_0} \right] \quad \dots(21)$$

【0065】

ただし、信号分析情報 I_0 が子音であるか否かを表す情報である場合には、減衰係数 β は、現在のフレームの信号分析情報 I_0 が子音であることを表す場合には0より大きく1より小さい予め定めた値であり($0 < \beta < 1$)、現在のフレームの信号分析情報 I_0 が子音でないことを表す場合には1である($\beta = 1$)。

【0066】

また、現在のフレームの信号分析情報 I_0 が子音らしさの指標値である場合には、減衰係数 β は、現在のフレームの信号分析情報 I_0 に基づいて決まる値であり、子音らしさの指標値 I_0 が大きいほど小さな値である。より具体的には、例えば、減衰係数 β は、子音らしさの指標値 I_0 が大きいほど小さな値であり、かつ、子音らしさの指標値 I_0 がその指標値が取り得る最小値である場合には $\beta = 1$ となり、かつ、子音らしさの指標値 I_0 がその指標値が取り得る最大値である場合には $\beta = 0$ となるような、所定の関数 $\beta = f(I_0)$ により求まるものとすればよい。

【0067】

なお、式(21)の A は、下記の式(22)により求まる振幅補正係数である。

【0068】

【数8】

$$A = \sqrt{1 + B_0^2 \sigma_0^2 \gamma_0^2} \quad \dots(22)$$

【0069】

また、 B_0 は予め定めた値であり、例えば3/4である。

【0070】

式(21)のピッチ強調処理は、ピッチ周期だけではなくピッチ利得も考慮したピッチ成分を強調する処理であり、かつ、子音であるフレームのピッチ成分については子音でないフレームのピッチ成分よりも強調の度合いを落としてピッチ成分を強調する処理である。

【0071】

つまり、信号分析情報 I_0 が子音であるか否か(スペクトル包絡が平坦であるか否か)を

10

20

30

40

50

表す場合、ピッチ強調部 130 では、子音である（スペクトル包絡が平坦である）と判定されたフレーム（時間区間）については、そのフレーム中の各時刻 n について、そのフレームのピッチ周期に対応するサンプル数 T_0 だけ、時刻 n よりも過去の時刻 $n - T_0$ の信号 X_{n-T_0} と、そのフレームのピッチ利得 G_{n-T_0} と、所定の定数 B_0 と、0 より大きく 1 より小さい値 α と、を乗算した信号と、時刻 n の信号 X_n と、を加算した信号を含む信号を出力信号 X^{out} として得る。また、ピッチ強調部 130 では、子音でない（スペクトル包絡が平坦でない）と判定されたフレーム（時間区間）については、そのフレーム中の各時刻 n について、そのフレームのピッチ周期に対応するサンプル数 T_0 だけ、時刻 n よりも過去の時刻 $n - T_0$ の信号 X_{n-T_0} と、そのフレームのピッチ利得 G_{n-T_0} と、所定の定数 B_0 と、を乗算した信号 $(B_0 \cdot \alpha \cdot X_{n-T_0})$ （この信号は式(21)において $\alpha=1$ に対応する）と、時刻 n の信号 X_n と、を加算した信号 $(X_n + B_0 \cdot \alpha \cdot X_{n-T_0})$ を含む信号を出力信号 X^{out} として得る。

10

【0072】

また、信号分析情報 I_0 が子音らしさの指標値（スペクトル包絡の平坦度合いの指標値）である場合、ピッチ強調部 130 では、そのフレーム中の各時刻 n について、信号 X_n を含むフレームのピッチ周期に対応するサンプル数 T_0 だけ、時刻 n よりも過去の時刻 $n - T_0$ の信号 X_{n-T_0} と、そのフレームのピッチ利得 G_{n-T_0} とそのフレームが子音らしければ子音らしいほど（そのフレームのスペクトル包絡が平坦であればあるほど）小さい値 B_0 と、を乗算した信号 $(B_0 \cdot \alpha \cdot X_{n-T_0})$ と、時刻 n の信号 X_n と、を加算した信号 $(X_n + B_0 \cdot \alpha \cdot X_{n-T_0})$ を含む信号を出力信号 X^{out} として得る。

20

【0073】

このピッチ強調処理により、子音のフレームであっても違和感を低減し、また、子音のフレームとそれ以外のフレームとが頻繁に切り替わる場合であっても、フレーム間におけるピッチ成分の強調の度合いの変動による違和感を低減する効果を得ることができる。

【0074】

[ピッチ強調処理 (S130) の第1変形例]

次に、音声ピッチ強調装置が行うピッチ強調処理の第1変形例とこれに関連する処理について説明する。

【0075】

第1変形例の音声ピッチ強調装置は、更にピッチ情報記憶部 150 を備える。

【0076】

ピッチ強調部 130 は、ピッチ分析部 120 が出力したピッチ周期とピッチ利得、信号特徴分析部 170 が出力した信号分析情報、及び音声ピッチ強調装置に入力された現在のフレームの時間領域の音信号を受け取り、現在のフレームの音信号サンプル列に対し、現在のフレームのピッチ周期 T_0 に対応するピッチ成分と、過去のフレームのピッチ周期に対応するピッチ成分と、を強調して得た出力信号のサンプル列を出力する。その際、現在のフレームのピッチ周期 T_0 に対応するピッチ成分については、現在のフレームのピッチ利得 G_0 に基づく強調の度合いが子音のフレーム（スペクトル包絡が平坦であるフレーム）のほうが子音以外のフレーム（スペクトル包絡が平坦でないフレーム）よりも小さくなるように、強調する。なお、以下の説明において、現在のフレームからみて s 個前のフレーム（ s 個過去のフレーム）のピッチ周期及びピッチ利得をそれぞれ T_{-s} 及び G_{-s} と表記する。

40

【0077】

ピッチ情報記憶部 150 には、1つ前のフレームから s 個過去のフレームまでのピッチ周期 T_{-1}, \dots, T_{-s} とピッチ利得 G_{-1}, \dots, G_{-s} とを記憶しておく。ただし、 s は予め定められた正の整数であり、例えば 1 である。

【0078】

ピッチ強調部 130 は、入力された現在のフレームのピッチ利得 G_0 と、ピッチ情報記憶部 150 から読み出した s 個過去のフレームのピッチ利得 G_{-s} と、入力された現在のフレームのピッチ周期 T_0 と、ピッチ情報記憶部 150 から読み出した s 個過去のフレームのピッチ周期 T_{-s} と、入力された現在のフレームの信号分析情報 I_0 とを用い、現在のフ

50

レームの音信号のサンプル列に対するピッチ強調処理を行う。

【 0 0 7 9 】

以下、具体例を説明する。

(ピッチ強調処理の第1変形例の具体例1)

この具体例では、ピッチ強調部130は、入力された現在のフレームの音信号のサンプル列を構成する各サンプル X_n ($L - N \leq n < L - 1$) に対して、以下の式(23)により出力信号 X_n^{new} を得ることにより、N個のサンプル X_{n-L}^{new} , ..., X_{n-1}^{new} による現在のフレームの出力信号のサンプル列を得る。

【 0 0 8 0 】

【数9】

$$X_n^{new} = \frac{1}{A} \left[X_n + B_0 \sigma_0 \gamma_0 X_{n-T_0} + B_{-\alpha} \sigma_{-\alpha} X_{n-T_{-\alpha}} \right] \quad \dots(23)$$

10

【 0 0 8 1 】

ただし、信号分析情報 l_i が子音であるか否かを表す情報である場合には、減衰係数 σ_i は、現在のフレームの信号分析情報 l_i が子音であることを表す場合には0より大きく1より小さい予め定めた値であり ($0 < \sigma_i < 1$)、現在のフレームの信号分析情報 l_i が子音でないことを表す場合には1である ($\sigma_i = 1$)。

【 0 0 8 2 】

また、現在のフレームの信号分析情報 l_i が子音らしさの指標値である場合には、減衰係数 σ_i は、現在のフレームの信号分析情報 l_i に基づいて決まる値であり、子音らしさの指標値 l_i が大きいほど小さな値である。より具体的には、例えば、減衰係数 σ_i は、子音らしさの指標値 l_i が大きいほど小さな値であり、かつ、子音らしさの指標値 l_i がその指標値が取り得る最小値である場合には $\sigma_i = 1$ となり、かつ、子音らしさの指標値 l_i がその指標値が取り得る最大値である場合には $\sigma_i = 0$ となるような、所定の関数 $\sigma_i = f(l_i)$ により求まるものとすればよい。

20

【 0 0 8 3 】

なお、式(23)のAは、下記の式(24)により求まる振幅補正係数である。

【 0 0 8 4 】

【数10】

$$A = \sqrt{1 + B_0^2 \sigma_0^2 \gamma_0^2 + B_{-\alpha}^2 \sigma_{-\alpha}^2 + 2B_0 B_{-\alpha} \sigma_0 \sigma_{-\alpha} \gamma_0} \quad \dots(24)$$

【 0 0 8 5 】

また、 B_0 と $B_{-\alpha}$ は、予め定めた1より小さい値であり、例えば3/4と1/4である。

【 0 0 8 6 】

(ピッチ強調処理の第1変形例の具体例2)

この具体例では、ピッチ強調部130は、入力された現在のフレームの音信号のサンプル列を構成する各サンプル X_n ($L - N \leq n < L - 1$) に対して、以下の式(25)により出力信号 X_n^{new} を得ることにより、N個のサンプル X_{n-L}^{new} , ..., X_{n-1}^{new} による現在のフレームの出力信号のサンプル列を得る。

【 0 0 8 7 】

【数11】

$$X_n^{new} = \frac{1}{A} \left[X_n + B_0 \sigma_0 \gamma_0 X_{n-T_0} + B_{-\alpha} \sigma_{-\alpha} \gamma_{-\alpha} X_{n-T_{-\alpha}} \right] \quad \dots(25)$$

40

【 0 0 8 8 】

ただし、減衰係数 σ_i は具体例1と同じであり、減衰係数 σ_{-i} は*i*個過去のフレームの減衰係数である。この具体例では*i*個過去のフレーム減衰係数 σ_{-i} を用いることから、この具体例の音声ピッチ強調装置は、更に減衰係数記憶部180を備える。減衰係数記憶部180には、1つ前のフレームから*i*個過去のフレームまでの減衰係数 $\sigma_{-1}, \dots, \sigma_{-i}$ を記憶しておく。

【 0 0 8 9 】

50

なお、式(25)のAは、下記の式(26)により求まる振幅補正係数である。

【 0 0 9 0 】

【 数 1 2 】

$$A = \sqrt{1 + B_0^2 \sigma_0^2 \gamma_0^2 + B_{-\alpha}^2 \sigma_{-\alpha}^2 \gamma_0^2 + 2B_0 B_{-\alpha} \sigma_0 \sigma_{-\alpha} \gamma_0 \gamma_{-\alpha}} \quad \dots(26)$$

【 0 0 9 1 】

また、B₀とB₋は、予め定めた1より小さい値であり、例えば3/4と1/4である。

【 0 0 9 2 】

(ピッチ強調処理の第1変形例の具体例3)

この具体例では、ピッチ強調部130は、入力された現在のフレームの音信号のサンプル列を構成する各サンプルX_n(L-N_n-L-1)に対して、以下の式(27)により出力信号X_n^{new}を得ることにより、N個のサンプルX_{L-1}^{new}, ..., X_{L-1}^{new}による現在のフレームの出力信号のサンプル列を得る。

10

【 0 0 9 3 】

【 数 1 3 】

$$X_n^{new} = \frac{1}{A} \left[X_n + B_0 \sigma_0 \gamma_0 X_{n-T_0} + B_{-\alpha} \sigma_{-\alpha} \gamma_0 X_{n-T_{-\alpha}} \right] \quad \dots(27)$$

【 0 0 9 4 】

ただし、減衰係数σ₀は具体例1や2と同じである。

【 0 0 9 5 】

また、式(27)のAは、下記の式(28)により求まる振幅補正係数である。

20

【 0 0 9 6 】

【 数 1 4 】

$$A = \sqrt{1 + B_0^2 \sigma_0^2 \gamma_0^2 + B_{-\alpha}^2 \sigma_{-\alpha}^2 \gamma_0^2 + 2B_0 B_{-\alpha} \sigma_0 \sigma_{-\alpha} \gamma_0 \gamma_{-\alpha}} \quad \dots(28)$$

【 0 0 9 7 】

また、B₀とB₋は、予め定めた1より小さい値であり、例えば3/4と1/4である。

【 0 0 9 8 】

この具体例は、具体例2の1個過去のフレームの減衰係数σ₀の代わりに現在のフレームの減衰係数σ₀を用いる構成である。この構成とすることにより、音声ピッチ強調装置が減衰係数記憶部180を備えずに済むようにできる。

30

【 0 0 9 9 】

第1変形例のピッチ強調処理は、ピッチ周期だけではなくピッチ利得も考慮したピッチ成分を強調する処理であり、かつ、子音であるフレームのピッチ成分については子音でないフレームのピッチ成分よりも強調の度合いを落としてピッチ成分を強調する処理であり、かつ、現在のフレームのピッチ周期T₀に対応するピッチ成分を強調しつつ、そのピッチ成分より少し強調の度合いを落として過去のフレームでのピッチ周期T₋に対応するピッチ成分も強調する処理である。第1変形例のピッチ強調処理により、短い時間区間(フレーム)ごとにピッチ強調処理を施す場合であっても、フレーム間におけるピッチ周期の変動による不連続性を低減する効果も得ることができる。

40

【 0 1 0 0 】

なお、信号分析情報I₀が子音であるか否かを表す情報である場合には、式(23)においてはB₀ > B₋とするのが好ましく、式(25)においてはB₀ > B₋とするのが好ましく、式(27)においてはB₀ > B₋とするのが好ましいが、式(23)においてB₀ = B₋としたり、式(25)においてB₀ = B₋としたり、式(27)においてB₀ = B₋としても、フレーム間におけるピッチ周期の変動による不連続性を低減する効果は奏される。

【 0 1 0 1 】

また、信号分析情報I₀が子音らしさの指標値である場合には、式(23)、式(25)、式(27)においてはB₀ > B₋とするのが好ましいが、B₀ = B₋としてもフレーム間におけるピッ

50

チ周期の変動による不連続性を低減する効果は奏される。

【 0 1 0 2 】

また、式(24)と式(26)と式(28)により求まる振幅補正係数Aは、現在のフレームのピッチ周期T₀と 個過去のフレームのピッチ周期T₋₁ とが十分に近い値であると仮定したときに、ピッチ成分のエネルギーがピッチ強調前後で保存されるようにするものである。

【 0 1 0 3 】

なお、ピッチ情報記憶部 1 5 0 は、現在のフレームのピッチ周期とピッチ利得を、以降のフレームのピッチ強調部 1 3 0 の処理において過去のフレームのピッチ周期とピッチ利得として用いることができるように、記憶内容を更新する。

【 0 1 0 4 】

また、減衰係数記憶部 1 8 0 を備える場合には、現在のフレームの減衰係数を、以降のフレームのピッチ強調部 1 3 0 の処理において過去のフレームの減衰係数として用いることができるように、記憶内容を更新する。

【 0 1 0 5 】

[ピッチ強調処理 (S 1 3 0) の第 2 変形例]

第 1 変形例では、現在のフレームの音信号サンプル列に対し、現在のフレームのピッチ周期T₀に対応するピッチ成分と、過去の 1 つのフレームのピッチ周期に対応するピッチ成分と、を強調して出力信号のサンプル列を得たが、過去の複数 (2 つ以上) のフレームのピッチ周期に対応するピッチ成分を強調するようにしてもよい。以下では、過去の複数のフレームのピッチ周期に対応するピッチ成分を強調する一例として、過去の 2 つのフレームのピッチ周期に対応するピッチ成分を強調する例について、第 1 変形例と異なる点を説明する。

【 0 1 0 6 】

ピッチ情報記憶部 1 5 0 には、現在のフレームより 個過去のフレームまでのピッチ周期T₋₁, ..., T₋₂, ..., T_{-M} とピッチ利得 G₋₁, ..., G₋₂, ..., G_{-M} とを記憶しておく。ただし、M は、 より大きい予め定めた正の整数である。例えば、M は 1 であり、M は 2 である。

【 0 1 0 7 】

ピッチ強調部 1 3 0 は、入力された現在のフレームのピッチ利得 G₀ と、ピッチ情報記憶部 1 5 0 から読み出した 個過去のフレームのピッチ利得 G₋₁ と、ピッチ情報記憶部 1 5 0 から読み出した 個過去のフレームのピッチ利得 G₋₂ と、入力された現在のフレームのピッチ周期T₀と、ピッチ情報記憶部 1 5 0 から読み出した 個過去のフレームのピッチ周期T₋₁ と、ピッチ情報記憶部 1 5 0 から読み出した 個過去のフレームのピッチ周期T₋₂ と、入力された現在のフレームの信号分析情報I₀とを用い、現在のフレームの音信号のサンプル列に対するピッチ強調処理を行う。

【 0 1 0 8 】

以下、具体例を説明する。

(ピッチ強調処理の第 2 変形例の具体例 1)

この具体例では、ピッチ強調部 1 3 0 は、入力された現在のフレームの音信号のサンプル列を構成する各サンプルX_n (L - N ≤ n ≤ L - 1) に対して、以下の式(29)により出力信号X_n^{new}を得ることにより、N個のサンプルX_{n-1}^{new}, ..., X_{n-N}^{new}による現在のフレームの出力信号のサンプル列を得る。

【 0 1 0 9 】

【 数 1 5 】

$$X_n^{new} = \frac{1}{A} \left[X_n + B_0 \sigma_0 \gamma_0 X_{n-T_0} + B_{-\alpha} \sigma_{-\alpha} X_{n-T_{-\alpha}} + B_{-\beta} \sigma_{-\beta} X_{n-T_{-\beta}} \right] \dots(29)$$

【 0 1 1 0 】

ただし、信号分析情報I₀が子音であるか否かを表す情報である場合には、減衰係数 G₀

10

20

30

40

50

は、現在のフレームの信号分析情報 I_i が子音であることを表す場合には0より大きく1より小さい予め定めた値であり ($0 < \gamma_i < 1$)、現在のフレームの信号分析情報 I_i が子音でないことを表す場合には1である ($\gamma_i = 1$)。

【 0 1 1 1 】

また、現在のフレームの信号分析情報 I_i が子音らしさの指標値である場合には、減衰係数 γ_i は、現在のフレームの信号分析情報 I_i に基づいて決まる値であり、子音らしさの指標値 I_i が大きいほど小さな値である。より具体的には、例えば、減衰係数 γ_i は、子音らしさの指標値 I_i が大きいほど小さな値であり、かつ、子音らしさの指標値 I_i がその指標値が取り得る最小値である場合には $\gamma_i = 1$ となり、かつ、子音らしさの指標値 I_i がその指標値が取り得る最大値である場合には $\gamma_i = 0$ となるような、所定の関数 $\gamma_i = f(I_i)$ により求めるものとすればよい。

10

【 0 1 1 2 】

なお、式(29)のAは、下記の式(30)により求まる振幅補正係数である。

【 0 1 1 3 】

【 数 1 6 】

$$A = \sqrt{1 + B_0^2 \sigma_0^2 \gamma_0^2 + B_{-\alpha}^2 \sigma_{-\alpha}^2 + B_{-\beta}^2 \sigma_{-\beta}^2 + E + F + G} \quad \dots(30)$$

ただし、

$$E = 2B_0 B_{-\alpha} \sigma_0 \sigma_{-\alpha} \gamma_0$$

$$F = 2B_0 B_{-\beta} \sigma_0 \sigma_{-\beta} \gamma_0$$

$$G = 2B_{-\alpha} B_{-\beta} \sigma_{-\alpha} \sigma_{-\beta}$$

【 0 1 1 4 】

また、 B_0 と $B_{-\alpha}$ と $B_{-\beta}$ は、予め定めた1より小さい値であり、例えば3/4と3/16と1/16である。

【 0 1 1 5 】

(ピッチ強調処理の第2変形例の具体例2)

この具体例では、ピッチ強調部130は、入力された現在のフレームの音信号のサンプル列を構成する各サンプル X_n ($L - N \leq n \leq L - 1$)に対して、以下の式(31)により出力信号 X_n^{new} を得ることにより、N個のサンプル X_{n-L}^{new} , ..., X_{n-1}^{new} による現在のフレームの出力信号のサンプル列を得る。

30

【 0 1 1 6 】

【 数 1 7 】

$$X_n^{new} = \frac{1}{A} \left[X_n + B_0 \sigma_0 \gamma_0 X_{n-T_0} + B_{-\alpha} \sigma_{-\alpha} \gamma_{-\alpha} X_{n-T_{-\alpha}} + B_{-\beta} \sigma_{-\beta} \gamma_{-\beta} X_{n-T_{-\beta}} \right] \quad \dots(31)$$

【 0 1 1 7 】

ただし、減衰係数 γ_i は具体例1と同じであり、減衰係数 γ_{-1} は1個過去のフレームの減衰係数であり、減衰係数 γ_{-2} は2個過去のフレームの減衰係数である。この具体例では1個過去のフレーム減衰係数 γ_{-1} と2個過去のフレーム減衰係数 γ_{-2} を用いることから、この具体例の音声ピッチ強調装置は、更に減衰係数記憶部180を備える。減衰係数記憶部180には、1つ前のフレームから1個過去のフレームまでの減衰係数 γ_{-1} , ..., γ_{-L} を記憶しておく。

40

【 0 1 1 8 】

なお、式(31)のAは、下記の式(32)により求まる振幅補正係数である。

【 0 1 1 9 】

【数 1 8】

$$A = \sqrt{1 + B_0^2 \sigma_0^2 \gamma_0^2 + B_{-\alpha}^2 \sigma_{-\alpha}^2 \gamma_{-\alpha}^2 + B_{-\beta}^2 \sigma_{-\beta}^2 \gamma_{-\beta}^2 + E + F + G} \quad \dots(32)$$

ただし、

$$E = 2B_0 B_{-\alpha} \sigma_0 \sigma_{-\alpha} \gamma_0 \gamma_{-\alpha}$$

$$F = 2B_0 B_{-\beta} \sigma_0 \sigma_{-\beta} \gamma_0 \gamma_{-\beta}$$

$$G = 2B_{-\alpha} B_{-\beta} \sigma_{-\alpha} \sigma_{-\beta} \gamma_{-\alpha} \gamma_{-\beta}$$

【0 1 2 0】

また、 B_0 と $B_{-\alpha}$ と $B_{-\beta}$ は、予め定めた1より小さい値であり、例えば3/4と3/16と1/16である。

【0 1 2 1】

(ピッチ強調処理の第2変形例の具体例3)

この具体例では、ピッチ強調部130は、入力された現在のフレームの音信号のサンプル列を構成する各サンプル X_n ($L - N \leq n \leq L - 1$)に対して、以下の式(33)により出力信号 X_n^{new} を得ることにより、N個のサンプル X_{n-L}^{new} 、 X_{n-L+1}^{new} による現在のフレームの出力信号のサンプル列を得る。

【0 1 2 2】

【数 1 9】

$$X_n^{new} = \frac{1}{A} \left[X_n + B_0 \sigma_0 \gamma_0 X_{n-T_0} + B_{-\alpha} \sigma_{-\alpha} \gamma_{-\alpha} X_{n-T_{-\alpha}} + B_{-\beta} \sigma_{-\beta} \gamma_{-\beta} X_{n-T_{-\beta}} \right] \quad \dots(33)$$

【0 1 2 3】

ただし、減衰係数 σ は具体例1や2と同じである。

【0 1 2 4】

また、式(33)のAは、下記の式(34)により求まる振幅補正係数である。

【0 1 2 5】

【数 2 0】

$$A = \sqrt{1 + B_0^2 \sigma_0^2 \gamma_0^2 + B_{-\alpha}^2 \sigma_{-\alpha}^2 \gamma_{-\alpha}^2 + B_{-\beta}^2 \sigma_{-\beta}^2 \gamma_{-\beta}^2 + E + F + G} \quad \dots(34)$$

ただし、

$$E = 2B_0 B_{-\alpha} \sigma_0 \sigma_{-\alpha} \gamma_0^2$$

$$F = 2B_0 B_{-\beta} \sigma_0 \sigma_{-\beta} \gamma_0^2$$

$$G = 2B_{-\alpha} B_{-\beta} \sigma_{-\alpha} \sigma_{-\beta} \gamma_0^2$$

【0 1 2 6】

また、 B_0 と $B_{-\alpha}$ と $B_{-\beta}$ は、予め定めた1より小さい値であり、例えば3/4と3/16と1/16である。

【0 1 2 7】

この具体例は、具体例2の $\sigma_{-\alpha}$ 個過去のフレームの減衰係数 $\sigma_{-\alpha}$ と $\sigma_{-\beta}$ 個過去のフレームの減衰係数 $\sigma_{-\beta}$ の代わりに現在のフレームの減衰係数 σ_0 を用いる構成である。この構成とすることにより、音声ピッチ強調装置が減衰係数記憶部180を備えずに済むようにできる。

【0 1 2 8】

第2変形例のピッチ強調処理も、第1変形例のピッチ強調処理と同様に、ピッチ周期だけではなくピッチ利得も考慮したピッチ成分を強調する処理であり、かつ、子音であるフレームのピッチ成分については子音でないフレームのピッチ成分よりも強調の度合いを落

としてピッチ成分を強調する処理であり、かつ、現在のフレームのピッチ周期 T_i に対応するピッチ成分を強調しつつ、そのピッチ成分より少し強調の度合いを落として過去のフレームでのピッチ周期に対応するピッチ成分も強調する処理である。第2変形例のピッチ強調処理により、短い時間区間(フレーム)ごとにピッチ強調処理を施す場合であっても、フレーム間におけるピッチ周期の変動による不連続性を低減する効果も得ることができる。

【0129】

なお、信号分析情報 I_i が子音であるか否かを表す情報である場合には、式(29)においては $B_{i+1} > B_i > B_{i-1}$ とするのが好ましく、式(31)においては $B_{i+1} > B_i > B_{i-1}$ とするのが好ましく、式(33)においては $B_i > B_{i-1} > B_{i-2}$ とするのが好ましいが、式(29)において $B_{i+1} > B_i$ や $B_{i+1} > B_{i-1}$ や $B_i > B_{i-1}$ としたり、式(31)において $B_{i+1} > B_{i-1}$ や $B_{i+1} > B_{i-2}$ や $B_i > B_{i-2}$ としたり、式(33)において $B_i > B_{i-2}$ や $B_i > B_{i-3}$ や $B_{i-1} > B_{i-3}$ としたり、式(33)においても、フレーム間におけるピッチ周期の変動による不連続性を低減する効果は奏される。

10

【0130】

また、信号分析情報 I_i が子音らしさの指標値である場合には、式(29)、式(31)、式(33)においては $B_i > B_{i-1} > B_{i-2}$ とするのが好ましいが、この大小関係を満たさなくともフレーム間におけるピッチ周期の変動による不連続性を低減する効果は奏される。

【0131】

また、式(30)と式(32)と式(34)により求まる振幅補正係数 A は、現在のフレームのピッチ周期 T_i と 個過去のフレームのピッチ周期 T_{i-1} と 個過去のフレームのピッチ周期 T_{i-2} とが十分に近い値であると仮定したときに、ピッチ成分のエネルギーがピッチ強調前後で保存されるようにするものである。

20

【0132】

(ピッチ強調処理のその他の変形例)

なお、振幅補正係数 A は、式(22)や式(24)や式(26)や式(28)や式(30)や式(32)や式(34)により求まる値ではなく、予め定めた1以上の値を用いてもよい。振幅補正係数 A を1とする場合には、ピッチ強調部130は、上記の式中の $1/A$ の項を含まないようにした式により出力信号 X^{out} を得るようにしてもよい。

【0133】

また、入力された音信号の各サンプルに加算する各ピッチ周期分前のサンプルに基づく値に代えて、例えばローパスフィルタを通した音信号における各ピッチ周期分前のサンプルを用いてもよいし、ローパスフィルタと等価な処理を行ってもよい。

30

【0134】

また、ピッチ利得が所定の閾値より小さい場合には、そのピッチ成分を含まないピッチ強調処理を行うようにしてもよい。例えば、現在のフレームのピッチ利得 G_i が所定の閾値より小さい場合には、現在のフレームのピッチ周期 T_i に対応するピッチ成分を出力信号に含めず、過去のフレームのピッチ利得が所定の閾値より小さい場合には、その過去のフレームのピッチ周期に対応するピッチ成分を出力信号に含めない構成としてもよい。

【0135】

また、信号特徴分析部170において子音らしさの指標値を得、信号分析情報 I_i としてピッチ強調部130に出力し、ピッチ強調部130において、子音らしさの指標値と閾値との大小関係に基づき強調度合い(減衰係数 α の大きさ)を2段階で異ならせる構成としてもよい。

40

【0136】

<第二実施形態>

第一実施形態と異なる部分を中心に説明する。

【0137】

本実施形態では、第一実施形態で説明したスペクトル包絡の平坦度合いの指標値(子音らしさの指標値)とは異なる子音らしさの指標値を用いる。

50

【0138】

信号特徴分析処理 (S170) の内容が第一実施形態とは異なる。

【0139】

[信号特徴分析処理 (S170)]

信号特徴分析部 170 には、第一実施形態と同様に時間領域の音信号に由来する情報が入力される。

【0140】

信号特徴分析部 170 は、現在のフレームが子音であるか否かを表す情報、または、現在のフレームの子音らしさの指標値、を得て、信号分析情報 I_0 としてピッチ強調部 130 へ出力する。

【0141】

また、例えば、信号特徴分析部 170 には、所定の時間長のフレーム (時間区間) 単位で、現在のフレームのピッチ周期 T_0 から N 個過去のフレームのピッチ周期 T_{-N} までが入力される。この場合には、信号特徴分析部 170 は、現在のフレームのピッチ周期 T_0 から N 個過去のフレームのピッチ周期 T_{-N} までを用いて、現在のフレームが子音であるか否かを表す情報、または、現在のフレームの子音らしさの指標値、を得て、信号分析情報 I_0 としてピッチ強調部 130 へ出力する。すなわち、この場合は、「時間領域の音信号に由来する情報」は現在のフレームのピッチ周期 T_0 から N 個過去のフレームのピッチ周期 T_{-N}

(図 1 中、一点鎖線で示す)までである。この場合には、音声ピッチ強調装置は更にピッチ情報記憶部 150 を備えて、ピッチ情報記憶部 150 には 1 つ前のフレームから N 個過去のフレームまでのピッチ周期 T_{-1}, \dots, T_{-N} を記憶しておく。そして、信号特徴分析部 170 は、ピッチ分析部 120 から入力された現在のフレームのピッチ周期 T_0 と、ピッチ情報記憶部 150 から読み出した 1 個過去のフレームから N 個過去のフレームまでのピッチ周期 T_{-1}, \dots, T_{-N} と、を用いる。 N は、予め定めた正の整数である。なお、ピッチ情報記憶部 150 は、現在のフレームのピッチ周期を、以降のフレームの信号特徴分析部 170 の処理において過去のフレームのピッチ周期として用いることができるように、記憶内容を更新する。

【0142】

信号特徴分析部 170 は例えば下記の例 2 - 1 から例 2 - 5 の信号特徴分析処理により信号分析情報 I_0 を得る。

【0143】

(信号特徴分析処理の例 2 - 1 : 子音らしさの指標値を信号分析情報とする例その 1)

この例では、信号特徴分析部 170 は、入力された現在のフレームのピッチ周期 T_0 から N 個過去のフレームのピッチ周期 T_{-N} までを用いて、現在のフレームの子音らしさの指標値としてピッチ周期の不連続性が大きいほど大きくなる指標値 (便宜上「子音らしさの第 2 - 1 の指標値」ともいう) を得て、得た第 2 - 1 の指標値を信号分析情報 I_0 として出力する。

【0144】

信号特徴分析部 170 は、例えば、ピッチ分析部 120 から入力されたピッチ周期 T_0 とピッチ情報記憶部 150 に記憶された 1 個過去のフレームから N 個過去のフレームまでのピッチ周期 T_{-1}, \dots, T_{-N} とを用いて、第 2 - 1 の指標値 I_0 を式(41)により求める。

$$I_0 = (T_0 \cdot T_{-1} + T_1 \cdot T_{-2} + \dots + T_{-(N-1)} \cdot T_{-N}) / (T_{-N}) \quad (41)$$

母音の場合には、ピッチ周期に連続性があり、連続するピッチ周期間の差分が 0 に近い値となり、 I_0 の値も小さくなる傾向がある。一方、子音の場合には、ピッチ周期に連続性がなく、 I_0 の値が大きくなる傾向がある。そこで、この例では、この傾向に基づき、第 2 - 1 の指標値 I_0 を子音らしさの指標値として利用する。なお、 I_0 は、判定するための十分な情報を得ることができる程度に大きく、かつ、 $T_0 \sim T_{-N}$ に対応する時間区間に子音と母音とが混在しない程度に小さい値とすることが望ましい。

【0145】

(信号特徴分析処理の例 2 - 2 : 子音らしさの指標値を信号分析情報とする例その 2)

10

20

30

40

50

この例では、信号特徴分析部 170 は、入力されたN個の時間領域の音信号サンプルを含む最新のJ個の音信号サンプルによるサンプル列を用いて、現在のフレームの子音らしさの指標値として摩擦音らしさの指標値（便宜上、「子音らしさの第2 - 2の指標値」ともいう）を得て、得た第2 - 2の指標値を信号分析情報 I_2 として出力する。

【0146】

信号特徴分析部 170 は、例えば、入力されたN個の時間領域の音信号サンプルを含む最新のJ個の音信号サンプルによるサンプル列の零交差点数(参考文献3参照)を摩擦音らしさの指標値である子音らしさの第2 - 2の指標値として求める。

(参考文献3) L.R.ラビナー 他著、鈴木久喜 訳、「音声のデジタル信号処理(上)」、株式会社コロナ社、1983年、p.132 137

【0147】

また、信号特徴分析部 170 は、例えば、入力されたN個の時間領域の音信号サンプルを含む最新のJ個の音信号サンプルによるサンプル列を修正離散コサイン変換(MDCT)などで周波数スペクトル系列に変換し、周波数スペクトル系列のうちの低域側にあるサンプルの平均エネルギーに対する周波数スペクトル系列のうちの高域側にあるサンプルの平均エネルギーの比が大きいくほど大きくなる指標値を摩擦音らしさの指標値である子音らしさの第2 - 2の指標値として求める。

【0148】

前述の通り、子音は、摩擦音を含む(参考文献1、参考文献2参照)。そこでこの例では、摩擦音らしさの指標値を子音らしさの指標値として利用する。

【0149】

(信号特徴分析処理の例2 - 3: 複数の指標値を組み合わせた指標値を信号分析情報とする例)

この例では、信号特徴分析部 170 は、まず、入力された現在のフレームのピッチ周期 T_1 から 個過去のフレームのピッチ周期 T_2 までを用いて、例2 - 1と同じ方法により、現在のフレームの子音らしさの第2 - 1の指標値を得る(Step 231)。信号特徴分析部 170 は、また、入力されたN個の時間領域の音信号サンプルを含む最新のJ個の音信号サンプルによるサンプル列を用いて、例2 - 2と同じ方法により、現在のフレームの子音らしさの第2 - 2の指標値を得る(Step 232)。信号特徴分析部 170 は、さらに、Step 231で得た第2 - 1の指標値とStep 232で得た第2 - 2の指標値の重み付け加算などにより、第2 - 1の指標値が大きくなるほど大きな値になり、かつ、第2 - 2の指標値が大きくなるほど大きな値になる値を、現在のフレームの子音らしさの指標値（便宜上、「第2 - 3の指標値」ともいう）として得て、得た第2 - 3の指標値を信号分析情報 I_3 として出力する(Step 233)。

【0150】

前述の通り、第2 - 1の指標値も第2 - 2の指標値も子音らしさを表す指標である。この例では2つの指標値を組み合わせることでより柔軟に子音らしさの指標値を設定することができる。

【0151】

信号特徴分析処理の例2 - 1 ~ 例2 - 3では子音らしさの指標値を信号分析情報とする例を説明してきた。ここからは、子音であるか否かを表す情報を信号分析情報とする例を説明する。

【0152】

(信号特徴分析処理の例2 - 4: 子音であるか否かを表す情報を信号分析情報とする例その1)

この例では、信号特徴分析部 170 は、まず、例2 - 1から例2 - 3の何れかと同じ方法により、現在のフレームの子音らしさの第2 - 1 ~ 2 - 3の指標値の何れかを得る。信号特徴分析部 170 は、次に、得た第2 - 1 ~ 2 - 3の指標値の何れかが予め定めた閾値以上または閾値を超える場合には、現在のフレームが子音であることを表す情報（「第2 - 1の指標値」 ~ 「第2 - 3の指標値」に対応する「現在のフレームが子音であるか否か

10

20

30

40

50

を表す情報」をそれぞれ、便宜上、「第2-1の情報」～「第2-3の情報」ともいう)を信号分析情報 I_0 として出力し、そうでない場合には、現在のフレームが子音でないことを表す第2-1～第2-3の情報の何れかを信号分析情報 I_0 として出力する。

【0153】

(信号特徴分析処理の例2-5：子音であるか否かを表す情報を信号分析情報とする例その2)

この例では、信号特徴分析部170は、まず、例2-1と同じ方法により、現在のフレームの子音らしさの第2-1の指標値を得て(Step 251)、Step 51で得た第2-1の指標値が予め定めた閾値以上または閾値を超える場合には、現在のフレームが子音であることを表す第2-1の情報を得て、そうでない場合には、現在のフレームが子音でないことを表す第2-1の情報を得る(Step 252)。信号特徴分析部170は、また、例2-2と同じ方法により、現在のフレームの子音らしさの第2-2の指標値を得て(Step 253)、Step 253で得た第2-2の指標値が予め定めた閾値以上または閾値を超える場合には、現在のフレームが子音であることを表す第2の情報を得て、そうでない場合には、現在のフレームが子音でないことを表す第2-2の情報を得る(Step 254)。信号特徴分析部170は、さらに、Step 252で得た第2-1の情報が子音であることを表しかつStep 254で得た第2-2の情報が子音であることを表す場合には、現在のフレームが子音であることを表す情報(便宜上、「第2-4の情報」ともいう)を信号分析情報 I_0 として出力し、そうでない場合には、現在のフレームが子音でないことを表す第2-4の情報を信号分析情報 I_0 として出力する(Step 255)。

【0154】

なお、信号特徴分析部170は、上記のStep 255に代えて、Step 252で得た第2-1の情報が子音であることを表すかまたはStep 254で得た第2-2の情報が子音であることを表す場合には、現在のフレームが子音であることを表す第2-4の情報を信号分析情報 I_0 として出力し、そうでない場合には、現在のフレームが子音でないことを表す第2-4の情報を信号分析情報 I_0 として出力してもよい(Step 255')。

【0155】

このような処理により信号特徴分析部170は、子音らしさの指標値または子音であるか否かを表す情報を信号分析情報 I_0 として出力する。

【0156】

<ピッチ強調部130>

ピッチ強調部130におけるピッチ強調処理(S130)は、第一実施形態と同様である。

【0157】

つまり、本実施形態のピッチ強調部130は、信号分析情報 I_0 が子音であるか否かを表す場合、子音であると判定されたフレーム(時間区間)については、そのフレーム中の各時刻 n について、そのフレームのピッチ周期に対応するサンプル数 T_0 だけ、時刻 n よりも過去の時刻 $n - T_0$ の信号 X_{n-T_0} と、そのフレームのピッチ利得 G_0 と、所定の定数 B_0 と、0より大きく1より小さい値と、を乗算した信号と、時刻 n の信号 X_n と、を加算した信号を含む信号を出力信号 $X_n^{(PW)}$ として得る。また、ピッチ強調部130は、子音でないとして判定されたフレーム(時間区間)については、そのフレーム中の各時刻 n について、そのフレームのピッチ周期に対応するサンプル数 T_0 だけ、時刻 n よりも過去の時刻 $n - T_0$ の信号 X_{n-T_0} と、そのフレームのピッチ利得 G_0 と、所定の定数 B_0 と、を乗算した信号($B_0 \cdot G_0 \cdot X_{n-T_0}$) (この信号は式(21)において $G_0=1$ に対応する)と、時刻 n の信号 X_n と、を加算した信号($X_n + B_0 \cdot G_0 \cdot X_{n-T_0}$)を含む信号を出力信号 $X_n^{(PW)}$ として得る。

【0158】

また、ピッチ強調部130では、信号分析情報 I_0 が子音らしさの指標値である場合、そのフレーム中の各時刻 n について、信号 X_n を含むフレームのピッチ周期に対応するサンプル数 T_0 だけ、時刻 n よりも過去の時刻 $n - T_0$ の信号 X_{n-T_0} と、そのフレームのピッチ利得 G_0 と、そのフレームが子音らしければ子音らしいほど小さい値 $B_0 \cdot G_0$ と、を乗算した信号($B_0 \cdot G_0 \cdot X_{n-T_0}$)

、 X_{n+1} と、時刻nの信号 X_n と、を加算した信号 $(X_n+B_n) + X_{n+1}$ を含む信号を出力信号 X_{n+1}^{new} として得る。

【0159】

なお、第一実施形態の第1変形例や第2変形例と同様のピッチ強調処理を行う場合には、信号特徴分析処理(S170)とピッチ強調処理(S130)においてピッチ情報記憶部150を共用してもよい。第一実施形態の第1変形例や第2変形例と同様のピッチ強調処理を行う場合には、 $>$ であってもよいし、 $<$ であってもよいし、 $=$ として重複する部分を最大限共用するようにしてもよい。同様に、第一実施形態の第2変形例と同様のピッチ強調処理を行う場合には、 $>$ であってもよいし、 $<$ であってもよいし、 $=$ として重複する部分を最大限共用するようにしてもよい。

10

【0160】

<効果>

上述の構成により、第一実施形態と同様の効果を得ることができる。

【0161】

<第三実施形態>

第一実施形態と異なる部分を中心に説明する。

【0162】

本実施形態では、第一実施形態で説明したスペクトル包絡の平坦度合いの指標値に加えて、第二実施形態で説明した子音らしさの指標値も用いて、子音らしさの指標値または子音であるか否かを表す情報を得る。

20

【0163】

信号特徴分析処理(S170)の内容が第一実施形態とは異なる。以下では、便宜上、第一実施形態で説明したスペクトル包絡の平坦度合いの指標値である子音らしさの第1-1~第1-5の指標値の何れかを第1の指標値と呼び、第二実施形態で説明した子音らしさの第2-1~第2-3の指標値の何れかを子音らしさの第2の指標値と呼び、子音らしさの第1の指標値と子音らしさの第2の指標値を用いて信号特徴分析処理(S170)で得る子音らしさの指標値を子音らしさの第3の指標値と呼ぶ。

【0164】

[信号特徴分析処理(S170)]

信号特徴分析部170は、第一実施形態で説明したスペクトル包絡の平坦度合いの指標値と、第二実施形態で説明した子音らしさの指標値と、に基づいて、子音らしさの指標値または子音であるか否かを表す情報を得て、信号分析情報としてピッチ強調部130へ出力する。信号特徴分析部170は、例えば下記の例3-1から例3-4の信号特徴分析処理により信号分析情報 I_1 を得る。

30

【0165】

(信号特徴分析処理の例3-1:スペクトル包絡の平坦度合いの指標値(子音らしさの第1の指標値)と子音らしさの第2の指標値とを組み合わせた指標値を子音らしさの第3の指標値とし、第3の指標値自体を信号分析情報とする例)

この例では、信号特徴分析部170は、まず、第一実施形態で説明した例1-1から1-5の何れかと同じ方法により、現在のフレームのスペクトル包絡の平坦度合いの指標値(子音らしさの第1の指標値)を得る(Step311)。信号特徴分析部170は、また、第二実施形態で説明した例2-1から例2-3の何れかの方法により、現在のフレームの子音らしさの第2の指標値を得る(Step312)。信号特徴分析部170は、さらに、Step311で得たスペクトル包絡の平坦度合いの指標値(子音らしさの第1の指標値)とStep312で得た子音らしさの第2の指標値の重み付け加算などにより、スペクトル包絡の平坦度合いの指標値(子音らしさの第1の指標値)が大きくなるほど大きな値になり、かつ、子音らしさの第2の指標値が大きくなるほど大きな値になる値を、現在のフレームの子音らしさの第3の指標値として得て、得た子音らしさの第3の指標値を信号分析情報 I_1 として出力する(Step313)。

40

【0166】

50

(信号特徴分析処理の例 3 - 2 : スペクトル包絡の平坦度合いの指標値 (子音らしさの第 1 の指標値) と子音らしさの第 2 の指標値とを組み合わせた第 3 の指標値を閾値判定して得た情報を信号分析情報とする例)

この例では、信号特徴分析部 170 は、まず、例 3 - 1 と同じ方法により、現在のフレームの子音らしさの第 3 の指標値を得る (Step 321)。信号特徴分析部 170 は、次に、Step 321 で得た子音らしさの第 3 の指標値が予め定めた閾値以上または閾値を超える場合には、現在のフレームが子音であることを表す第 3 の情報を信号分析情報₁として出力し、そうでない場合には、現在のフレームが子音でないことを表す第 3 の情報を信号分析情報₁として出力する。

【0167】

(信号特徴分析処理の例 3 - 3 : 子音であるかまたはスペクトル包絡が平坦であるか否かを表す情報を信号分析情報とする例)

この例では、信号特徴分析部 170 は、まず、第一実施形態で説明した例 1 - 1 から例 1 - 5 の何れかと同じ方法により、現在のフレームのスペクトル包絡の平坦度合いの指標値 (子音らしさの第 1 の指標値) を得て (Step 331)、Step 331 で得た第 1 の指標値が予め定めた閾値以上または閾値を超える場合には、現在のフレームのスペクトル包絡が平坦である (現在のフレームが子音である) ことを表す第 1 の情報を得て、そうでない場合には、現在のフレームのスペクトル包絡が平坦でない (現在のフレームが子音でない) ことを表す第 1 の情報を得る (Step 332)。信号特徴分析部 170 は、また、第二実施形態で説明した例 2 - 1 から例 2 - 3 の何れかの方法により、子音らしさの第 2 の指標値を得て (Step 333)、Step 333 で得た第 2 の指標値が予め定めた閾値以上または閾値を超える場合には、現在のフレームが子音であることを表す第 2 の情報を得て、そうでない場合には、現在のフレームが子音でないことを表す第 2 の情報を得る (Step 334)。信号特徴分析部 170 は、さらに、Step 332 で得た第 1 の情報がスペクトル包絡が平坦である (子音である) ことを表すかまたは Step 334 で得た第 2 の情報が子音であることを表す場合には、現在のフレームが子音であることを表す第 3 の情報を信号分析情報₁として出力し、そうでない場合には、現在のフレームが子音でないことを表す第 3 の情報を信号分析情報₁として出力する。

【0168】

(信号特徴分析処理の例 3 - 4 : 子音でありかつスペクトル包絡が平坦であるか否かを表す情報を信号分析情報とする例)

この例では、信号特徴分析部 170 は、まず、第一実施形態で説明した例 1 - 1 から例 1 - 5 の何れかと同じ方法により、現在のフレームの子音らしさの第 1 の指標値を得て (Step 341)、Step 341 で得た指標値が予め定めた閾値以上または閾値を超える場合には、現在のフレームのスペクトル包絡が平坦である (現在のフレームが子音である) ことを表す第 1 の情報を得て、そうでない場合には、現在のフレームのスペクトル包絡が平坦でない (現在のフレームが子音でない) ことを表す第 1 の情報を得る (Step 342)。信号特徴分析部 170 は、また、第二実施形態で説明した例 2 - 1 から例 2 - 3 の何れかの方法により、現在のフレームの子音らしさの第 2 の指標値を得て (Step 343)、Step 343 で得た指標値が予め定めた閾値以上または閾値を超える場合には、現在のフレームが子音であることを表す第 2 の情報を得て、そうでない場合には、現在のフレームが子音でないことを表す第 2 の情報を得る (Step 344)。信号特徴分析部 170 は、さらに、Step 342 で得た第 1 の情報がスペクトル包絡が平坦である (子音である) ことを表しかつ Step 344 で得た第 2 の情報が子音であることを表す場合には、現在のフレームが子音であることを表す第 3 の情報を信号分析情報₁として出力し、そうでない場合には、現在のフレームが子音でないことを表す第 3 の情報を信号分析情報₁として出力する。

【0169】

<ピッチ強調部 130>

ピッチ強調部 130 におけるピッチ強調処理 (S130) は、第一実施形態と同様である。

10

20

30

40

50

【0170】

つまり、本実施形態のピッチ強調部130は、信号分析情報 I_0 が子音であるか否かを表す場合(第3の情報の場合)、信号 X_n のスペクトル包絡が平坦である、またはノおよび、子音であると判定されたフレーム(時間区間)については、そのフレームの各時刻 n について、そのフレームのピッチ周期に対応するサンプル数 T_0 だけ、時刻 n よりも過去の時刻 $n - T_0$ の信号 X_{n-T_0} と、そのフレームのピッチ利得 G_0 と、所定の定数 B_0 と、0より大きく1より小さい値と、を乗算した信号と、時刻 n の信号 X_n と、を加算した信号を含む信号を出力信号 X^{out} として得る。また、ピッチ強調部130は、それ以外の判定がされたフレームについては、そのフレームの各時刻 n について、そのフレームのピッチ周期に対応するサンプル数 T_0 だけ、時刻 n よりも過去の時刻 $n - T_0$ の信号 X_{n-T_0} と、そのフレームのピッチ利得 G_0 と、所定の定数 B_0 と、を乗算した信号($B_0 \cdot X_{n-T_0}$)(この信号は式(21)において $\alpha=1$ に対応する)と、時刻 n の信号 X_n と、を加算した信号($X_n + B_0 \cdot X_{n-T_0}$)を含む信号を出力信号 X^{out} として得る(例3-3、3-4に対応する)。なお、例3-2では、スペクトル包絡の平坦度合いの指標値(子音らしさの第1の指標値)と子音らしさの第2の指標値とを組み合わせた第3の指標値を閾値判定しており、この閾値判定は、信号 X_n のスペクトル包絡が平坦である、またはノおよび、子音であるか否かを判定することに相当する。

10

【0171】

また、ピッチ強調部130では、信号分析情報 I_0 が子音らしさの指標値である場合(第3の指標値の場合)、そのフレームの各時刻 n について、信号 X_n を含むフレームのピッチ周期に対応するサンプル数 T_0 だけ、時刻 n よりも過去の時刻 $n - T_0$ の信号 X_{n-T_0} と、そのフレームのピッチ利得 G_0 と、そのフレームのスペクトル包絡が平坦であればあるほど小さくかつそのフレームが子音らしければ子音らしいほど小さい値 B_0 と、を乗算した信号($B_0 \cdot X_{n-T_0}$)と、時刻 n の信号 X_n と、を加算した信号($X_n + B_0 \cdot X_{n-T_0}$)を含む信号を出力信号 X^{out} として得る(例3-1に対応する)。

20

【0172】

<効果>

このような構成とすることで、第一実施形態と同様の効果を得ることができる。さらに、本実施系形態では、第1の指標値(スペクトル包絡の平坦度合いの指標値)に加えて、第2の指標値も考慮することで、より適切な子音らしさの指標値を得ることができる。

【0173】

<その他の変形例>

音声ピッチ強調装置外で行われる復号処理などにより各フレームのピッチ周期とピッチ利得と信号分析情報を得られている場合には、音声ピッチ強調装置を図3の構成として、音声ピッチ強調装置外で得られたピッチ周期とピッチ利得と信号分析情報に基づきピッチを強調してもよい。図4はその処理フローを示す。この場合には、第一実施形態、第二実施形態、第三実施形態、およびそれらの変形例の音声ピッチ強調装置が備える自己相関関数算出部110やピッチ分析部120や信号特徴分析部170や自己相関関数記憶部160を備える必要はなく、ピッチ強調部130が、ピッチ分析部120が出力したピッチ周期とピッチ利得と信号特徴分析部170が出力した信号分析情報ではなく、音声ピッチ強調装置に入力されたピッチ周期とピッチ利得と信号分析情報を用いてピッチ強調処理(S130)を行うようにすればよい。このような構成とすれば、音声ピッチ強調装置自体の演算処理量は第一実施形態、第二実施形態、第三実施形態、およびそれらの変形例よりも少なくすることが可能である。ただし、第一実施形態、第二実施形態、第三実施形態、およびそれらの変形例の音声ピッチ強調装置は、音声ピッチ強調装置外のピッチ周期やピッチ利得と信号分析情報を得る頻度に依存せずピッチ周期やピッチ利得と信号分析情報を得ることができることから、非常に短い時間長のフレーム単位でのピッチ強調処理を行うことが可能である。上記のサンプリング周波数32kHzの例であれば、Nを例えば32とすれば、1msのフレーム単位でピッチ強調処理を行うことができる。

30

40

【0174】

なお、以上の説明では、音信号そのものに対してピッチ強調処理を施すことを前提とし

50

ていたが、非特許文献1に記載されているような線形予測残差に対してピッチ強調処理を行ってから線形予測合成をするような構成における、線形予測残差に対するピッチ強調処理として本発明を適用してもよい。すなわち、本発明を、音信号そのものではなく、音信号に対して分析や加工をして得た信号などの音信号に由来する信号に対して適用してもよい。

【0175】

本発明は上記の実施形態及び変形例に限定されるものではない。例えば、上述の各種の処理は、記載に従って時系列に実行されるのみならず、処理を実行する装置の処理能力あるいは必要に応じて並列的にあるいは個別に実行されてもよい。その他、本発明の趣旨を逸脱しない範囲で適宜変更が可能である。

【0176】

<プログラム及び記録媒体>

また、上記の実施形態及び変形例で説明した各装置における各種の処理機能をコンピュータによって実現してもよい。その場合、各装置が有すべき機能の処理内容はプログラムによって記述される。そして、このプログラムをコンピュータで実行することにより、上記各装置における各種の処理機能がコンピュータ上で実現される。

【0177】

この処理内容を記述したプログラムは、コンピュータで読み取り可能な記録媒体に記録しておくことができる。コンピュータで読み取り可能な記録媒体としては、例えば、磁気記録装置、光ディスク、光磁気記録媒体、半導体メモリ等のようなものでもよい。

【0178】

また、このプログラムの流通は、例えば、そのプログラムを記録したDVD、CD-ROM等の可搬型記録媒体を販売、譲渡、貸与等することによって行う。さらに、このプログラムをサーバコンピュータの記憶装置に格納しておき、ネットワークを介して、サーバコンピュータから他のコンピュータにそのプログラムを転送することにより、このプログラムを流通させてもよい。

【0179】

このようなプログラムを実行するコンピュータは、例えば、まず、可搬型記録媒体に記録されたプログラムもしくはサーバコンピュータから転送されたプログラムを、一旦、自己の記憶部に格納する。そして、処理の実行時、このコンピュータは、自己の記憶部に格納されたプログラムを読み取り、読み取ったプログラムに従った処理を実行する。また、このプログラムの別の実施形態として、コンピュータが可搬型記録媒体から直接プログラムを読み取り、そのプログラムに従った処理を実行することとしてもよい。さらに、このコンピュータにサーバコンピュータからプログラムが転送されるたびに、逐次、受け取ったプログラムに従った処理を実行することとしてもよい。また、サーバコンピュータから、このコンピュータへのプログラムの転送は行わず、その実行指示と結果取得のみによって処理機能を実現する、いわゆるASP(Application Service Provider)型のサービスによって、上述の処理を実行する構成としてもよい。なお、プログラムには、電子計算機による処理の用に供する情報であってプログラムに準ずるもの(コンピュータに対する直接の指令ではないがコンピュータの処理を規定する性質を有するデータ等)を含むものとする。

【0180】

また、コンピュータ上で所定のプログラムを実行させることにより、各装置を構成することとしたが、これらの処理内容の少なくとも一部をハードウェア的に実現することとしてもよい。

10

20

30

40

【図1】

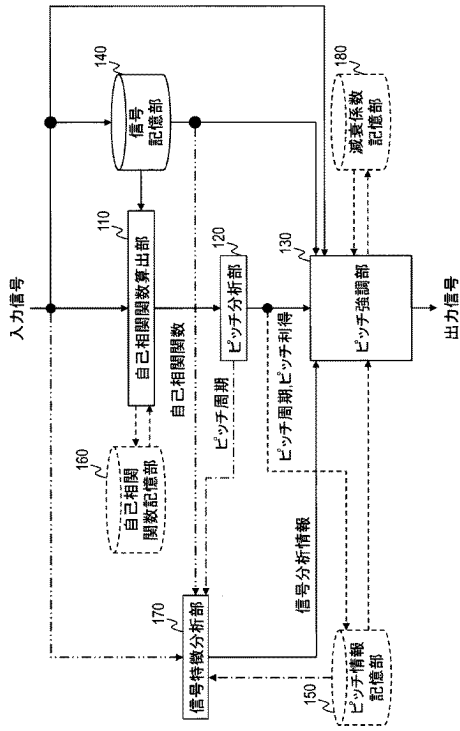


図1

【図2】

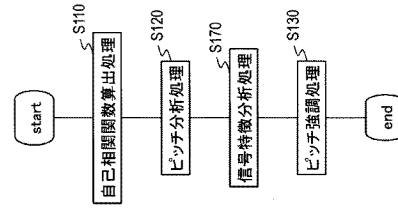


図2

【図3】

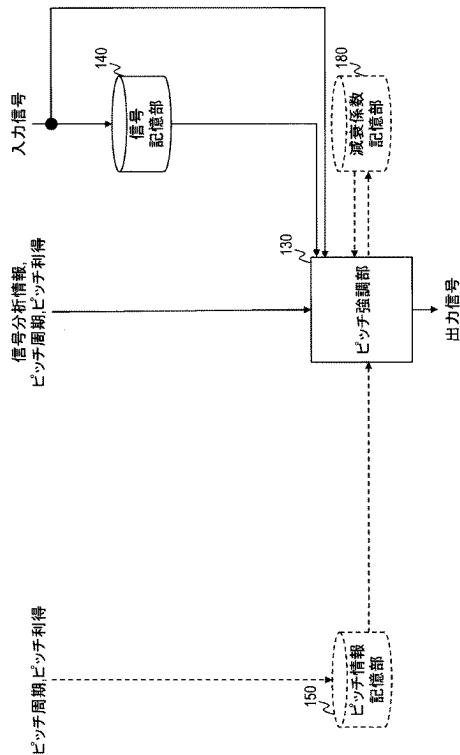


図3

【図4】

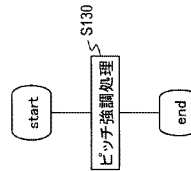


図4

フロントページの続き

(72)発明者 守谷 健弘

東京都千代田区大手町一丁目5番1号 日本電信電話株式会社内