

(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2019-197150
(P2019-197150A)

(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 16 頁)

(21) 出願番号	特願2018-91201 (P2018-91201)	(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、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、 B を1より大きい値とし、当該時間区間の各時刻 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_n 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 B の乗と、所定の定数 B_0 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。

【選択図】 図1

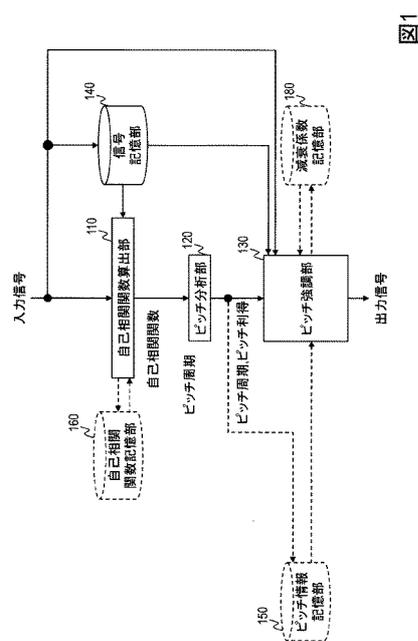


図1

【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、

を 1 より大きい値とし、当該時間区間の各時刻 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_n 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_n の乗と、所定の定数 B_n 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。

10

【請求項 2】

請求項 1 に記載のピッチ強調装置であって、

前記ピッチ強調部は、

当該時間区間の各時刻 n について、

前記加算した信号に、

当該時間区間よりも M 個過去の時間区間のピッチ周期に対応するサンプル数 T_{n-M} だけ当該時刻 n よりも過去の時刻の前記信号と、当該時間区間よりも M 個過去の時間区間のピッチ利得 G_{n-M} と、所定の定数 B_{n-M} と、を乗算した信号

も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ものである

ピッチ強調装置。

20

【請求項 3】

請求項 1 に記載のピッチ強調装置であって、

前記ピッチ強調部は、

当該時間区間の各時刻 n について、

前記加算した信号に、

当該時間区間よりも M 個過去の時間区間のピッチ周期に対応するサンプル数 T_{n-M} だけ当該時刻 n よりも過去の時刻の前記信号と、当該時間区間よりも M 個過去の時間区間のピッチ利得 G_{n-M} の乗と、所定の定数 B_{n-M} と、を乗算した信号

も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ものである

ピッチ強調装置。

30

【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、

を 1 より大きい値とし、当該時間区間の各時刻 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_n 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_n の乗と、所定の定数 B_n 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。

40

【請求項 5】

請求項 1 から請求項 3 の何れか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】

50

【 0 0 0 2 】

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）。

【 0 0 0 3 】

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 】

【 0 0 0 4 】

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

【 特許文献 】

【 0 0 0 5 】

【 特許文献 1 】 特開平 1 0 - 1 4 3 1 9 5 号公報

【 発明の概要 】

【 発明が解決しようとする課題 】

【 0 0 0 6 】

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。

【 0 0 0 7 】

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 】

【 0 0 0 8 】

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、 B を1より大きい値とし、当該時間区間の各時刻 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 の乗と、所定の定数 B_i と、 B_i を乗算した信号と、当該時刻 n の信号と、 B_i 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

30

40

50

【発明の効果】

【0009】

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。

【図面の簡単な説明】

【0010】

【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。

10

【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。

【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】

【0011】

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。

20

【0012】

< 第一実施形態 >

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。

【0013】

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、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声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、ピッチ成分にピッチ利得そのものではなく、ピッチ利得の乗を乗算する。ただし、 > 1 である。子音には母音に比べて周期性が小さいという性質があり、入力された信号を分析して得られるピッチ利得は、子音の時間区間のほうが母音の時間区間より小さな値となる。なお、このピッチの利得は、例外的な場合を除き、通常1より小さい値である。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、ピッチ成分にピッチ利得そのものではなく、ピッチ利得の乗を乗算することで、子音の時間区間のピッチ成分の強調の度合いを母音の時間区間よりも小さくする。

30

【0014】

第一実施形態の音声ピッチ強調装置は、自己相関関数算出部110とピッチ分析部120とピッチ強調部130と信号記憶部140とを備えるものであり、更にピッチ情報記憶部150と自己相関関数記憶部160と減衰係数記憶部180とを備えてもよい。

40

【0015】

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（R

50

andom 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。

【0016】

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(S110)とピッチ分析処理(S120)とピッチ強調処理(S130)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(S110)とピッチ分析処理(S120)とピッチ強調処理(S130)のそれぞれについて、関連する処理と共に説明する。

10

【0017】

[自己相関関数算出処理(S110)]

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。

【0018】

自己相関関数算出部110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110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。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。

20

【0019】

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。

30

40

【0020】

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

50

する。

【 0 0 2 1 】

【 数 1 】

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

【 0 0 2 2 】

自己相関関数算出部 1 1 0 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 1 2 0 に出力する。

【 0 0 2 3 】

なお、この時間差 $(1), \dots, (M)$ は後述するピッチ分析部 1 2 0 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 $(1), \dots, (M)$ 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 L で除算した正規化自己相関関数 R/L 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/L 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。

【 0 0 2 4 】

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 1 6 0 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） $R_{(1)}, \dots, R_{(M)}$ を記憶しておき、自己相関関数算出部 1 1 0 は、自己相関関数記憶部 1 6 0 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） $R_{(1)}, \dots, R_{(M)}$ 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 $R_{(1)}, \dots, R_{(M)}$ 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、 $(1), \dots, (M)$ のそれぞれを τ とすると、自己相関関数算出部 1 1 0 は、現在のフレームの自己相関関数 R を、直前のフレームの処理で得られた自己相関関数 R （直前のフレームの自己相関関数 R ）に対して、以下の式(2)で得られる差分 ΔR^+ を加算し、式(3)で得られる差分 ΔR^- を減算することにより得る。

【 0 0 2 5 】

【 数 2 】

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

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

【 0 0 2 6 】

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 $(1), \dots, (M)$ 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 $(1), \dots, (M)$ は、75 から 320 の約半分である 37 から 160 とすればよい。

10

20

30

50

【 0 0 2 7 】

信号記憶部 1 4 0 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 1 3 0 の処理までを終えた後に、その時点で最新の $L - N$ 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、 $L > 2N$ の場合、信号記憶部 1 4 0 は、記憶されている $L - N$ 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 X_0, X_1, \dots, X_{N-1} を削除し、 $X_N, X_{N+1}, \dots, X_{L-N-1}$ を $X_0, X_1, \dots, X_{L-2N-1}$ 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 $X_{L-2N}, X_{L-2N+1}, \dots, X_{L-N-1}$ として新たに記憶する。また、 $L = 2N$ の場合、信号記憶部 1 4 0 は、記憶されている $L - N$ 個の音信号サンプル $X_0, X_1, \dots, X_{L-N-1}$ 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 $L - N$ 個の音信号サンプルを $X_0, X_1, \dots, X_{L-N-1}$ として新たに記憶する。なお、 $L = N$ である場合には、音声ピッチ強調装置内には信号記憶部 1 4 0 を備える必要はない。

10

【 0 0 2 8 】

また、自己相関関数記憶部 1 6 0 は、自己相関関数算出部 1 1 0 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 $R_{(1)}, \dots, R_{(M)}$ 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 1 6 0 は、記憶されている $R_{(1)}, \dots, R_{(M)}$ を削除し、算出した現在のフレームの自己相関関数 $R_{(1)}, \dots, R_{(M)}$ を新たに記憶する。

【 0 0 2 9 】

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 1 1 0 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_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 $(1), \dots, (M)$ それぞれに対する自己相関関数 $R_{(1)}, \dots, R_{(M)}$ を算出すればよい。

20

【 0 0 3 0 】

[ピッチ分析処理 (S 1 2 0)]

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。

【 0 0 3 1 】

ピッチ分析部 1 2 0 には、自己相関関数算出部 1 1 0 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。

30

【 0 0 3 2 】

ピッチ分析部 1 2 0 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 $R_{(1)}, \dots, R_{(M)}$ の中で最大の値を求め、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 ρ 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_0 として得て、それぞれをピッチ強調部 1 3 0 へ出力する。

【 0 0 3 3 】

[ピッチ強調処理 (S 1 3 0)]

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。

40

【 0 0 3 4 】

ピッチ強調部 1 3 0 は、ピッチ分析部 1 2 0 が出力したピッチ周期とピッチ利得、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 ρ の乗 (> 1) に比例した強調の度合いで強調して得た出力信号のサンプル列を出力する。

【 0 0 3 5 】

以下、具体例を説明する。

【 0 0 3 6 】

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 ρ と、入力された現

50

在のフレームのピッチ周期 T_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(4)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} , ..., 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。

【0037】

【数3】

$$X_n^{new} = \frac{1}{A} [X_n + B_0 \sigma_0^{2\eta} X_{n-T_0}] \quad \dots(4)$$

【0038】

ただし、 A は1より大きい所定の値である。なお、式(4)の A は、下記の式(5)により求まる振幅補正係数である。

【0039】

【数4】

$$A = \sqrt{1 + B_0^2 \sigma_0^{2\eta}} \quad \dots(5)$$

【0040】

また、 B_0 は予め定めた値であり、例えば3/4である。ピッチ利得 σ_0 は、例外的な場合を除き、通常は1より小さい値である。また、例外的に1より大きな値がピッチ利得 σ_0 として求まってしまった場合には、ピッチ利得 σ_0 を1に置き換えてから上記式(4)のピッチ強調処理を行えばよい。従って、式(4)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、ピッチ利得が小さいフレームのピッチ成分についてはピッチ利得が大き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。

【0041】

つまり、ピッチ強調部130では、フレーム(時間区間)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 σ_0 の乗 $\sigma_0^{2\eta}$ と、所定の定数 B_0 と、を乗算した信号($B_0 \sigma_0^{2\eta} X_{n-T_0}$)と、時刻 n の信号 X_n と、を加算した信号($X_n + B_0 \sigma_0^{2\eta} X_{n-T_0}$)を含む信号を出力信号 X_n^{new} として得る。

【0042】

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。

【0043】

[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。

【0044】

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。

【0045】

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、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 σ_0 の乗($\sigma_0 > 1$)に比例した強調の度合いで、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_{0s} 及び σ_{0s} と表記する。

【0046】

10

20

30

40

50

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

【0047】

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 と、ピッチ情報記憶部 150 から読み出した 個過去のフレームのピッチ利得 と、入力された現在のフレームのピッチ周期 T_t と、ピッチ情報記憶部 150 から読み出した 個過去のフレームのピッチ周期 T_{-} 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。

【0048】

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)

具体例1は、現在のフレームのピッチ周期 T_t 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 の乗 (> 1) に比例した強調の度合いで強調し、 個過去のフレームのピッチ周期 T_{-} に対応するピッチ成分については、 個過去のフレームのピッチ利得 に比例した強調の度合いで強調する例である。

【0049】

すなわち、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(6)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 $X_{n-N}^{old}, \dots, X_{n-1}^{old}$ による現在のフレームの出力信号のサンプル列を得る。

【0050】

【数5】

$$X_n^{new} = \frac{1}{A} \left[X_n + B_0 \sigma_0^\eta X_{n-T_0} + B_{-\alpha} \sigma_{-\alpha} X_{n-T_{-\alpha}} \right] \quad \dots(6)$$

【0051】

なお、式(6)のAは、下記の式(7)により求まる振幅補正係数である。

【0052】

【数6】

$$A = \sqrt{1 + B_0^2 \sigma_0^{2\eta} + B_{-\alpha}^2 \sigma_{-\alpha}^2 + 2B_0 B_{-\alpha} \sigma_0^\eta \sigma_{-\alpha}} \quad \dots(7)$$

【0053】

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。

【0054】

(ピッチ強調処理の第1変形例の具体例2)

具体例2は、現在のフレームのピッチ周期 T_t 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 の乗 (> 1) に比例した強調の度合いで強調し、 個過去のフレームのピッチ周期 T_{-} に対応するピッチ成分については、 個過去のフレームのピッチ利得 の乗に比例した強調の度合いで強調する例である。

【0055】

すなわち、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(8)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 $X_{n-N}^{old}, \dots, X_{n-1}^{old}$ による現在のフレームの出力信号のサンプル列を得る。

【0056】

【数7】

$$X_n^{new} = \frac{1}{A} \left[X_n + B_0 \sigma_0^\eta X_{n-T_0} + B_{-\alpha} \sigma_{-\alpha}^\eta X_{n-T_{-\alpha}} \right] \quad \dots(8)$$

【0057】

なお、式(8)のAは、下記の式(9)により求まる振幅補正係数である。

【 0 0 5 8 】

【 数 8 】

$$A = \sqrt{1 + B_0^2 \sigma_0^{2\eta} + B_{-\alpha}^2 \sigma_{-\alpha}^{2\eta} + 2B_0 B_{-\alpha} \sigma_0^\eta \sigma_{-\alpha}^\eta} \quad \dots(9)$$

【 0 0 5 9 】

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 6 0 】

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、ピッチ利得が小さいフレームのピッチ成分についてはピッチ利得が大き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_0 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_{-1} 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。

10

【 0 0 6 1 】

なお、式(6),(8)においては $B_0 > B_{-\alpha}$ とするのが好ましいが、式(6),(8)において $B_0 = B_{-\alpha}$ 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 0 6 2 】

また、式(7)と式(9)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_0 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。

20

【 0 0 6 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 0 6 4 】

[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 T_0 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。

30

【 0 0 6 5 】

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 $T_{-1}, \dots, T_{-2}, \dots, T_{-N}$ とピッチ利得 $G_{-1}, \dots, G_{-2}, \dots, G_{-N}$ とを記憶しておく。ただし、 N は、 1 より大きい予め定めた正の整数である。例えば、 N は 1 であり、 N は 2 である。

40

【 0 0 6 6 】

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 G_0 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 T_0 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 0 6 7 】

50

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)

具体例1は、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 の乗(> 1)に比例した強調の度合いで強調し、 β 個過去のフレームのピッチ周期 $T_{-\beta}$ に対応するピッチ成分については、 β 個過去のフレームのピッチ利得 $G_{-\beta}$ に比例した強調の度合いで強調し、 α 個過去のフレームのピッチ周期 $T_{-\alpha}$ に対応するピッチ成分については、 α 個過去のフレームのピッチ利得 $G_{-\alpha}$ に比例した強調の度合いで強調する例である。

【0068】

すなわち、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(10)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} , ..., 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。

【0069】

【数9】

$$X_n^{new} = \frac{1}{A} \left[X_n + B_0 \sigma_0^\eta X_{n-T_0} + B_{-\alpha} \sigma_{-\alpha} X_{n-T_{-\alpha}} + B_{-\beta} \sigma_{-\beta} X_{n-T_{-\beta}} \right] \quad \dots(10)$$

【0070】

なお、式(10)の A は、下記の式(11)により求まる振幅補正係数である。

【0071】

【数10】

$$A = \sqrt{1 + B_0^2 \sigma_0^{2\eta} + B_{-\alpha}^2 \sigma_{-\alpha}^2 + B_{-\beta}^2 \sigma_{-\beta}^2 + E + F + G} \quad \dots(11)$$

ただし、

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

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

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

【0072】

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。

【0073】

(ピッチ強調処理の第2変形例の具体例2)

具体例2は、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 の乗(> 1)に比例した強調の度合いで強調し、 β 個過去のフレームのピッチ周期 $T_{-\beta}$ に対応するピッチ成分については、 β 個過去のフレームのピッチ利得 $G_{-\beta}$ の乗に比例した強調の度合いで強調し、 α 個過去のフレームのピッチ周期 $T_{-\alpha}$ に対応するピッチ成分については、 α 個過去のフレームのピッチ利得 $G_{-\alpha}$ の乗に比例した強調の度合いで強調する例である。

【0074】

すなわち、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(12)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} , ..., 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。

【0075】

10

20

40

【数 1 1】

$$X_n^{new} = \frac{1}{A} \left[X_n + B_0 \sigma_0^\eta X_{n-T_0} + B_{-\alpha} \sigma_{-\alpha}^\eta X_{n-T_{-\alpha}} + B_{-\beta} \sigma_{-\beta}^\eta X_{n-T_{-\beta}} \right] \quad \dots(12)$$

【0076】

なお、式(12)のAは、下記の式(13)により求まる振幅補正係数である。

【0077】

【数 1 2】

$$A = \sqrt{1 + B_0^2 \sigma_0^{2\eta} + B_{-\alpha}^2 \sigma_{-\alpha}^{2\eta} + B_{-\beta}^2 \sigma_{-\beta}^{2\eta} + E + F + G} \quad \dots(13)$$

ただし、

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

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

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

【0078】

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。

20

【0079】

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、子音のピッチ利得が小さいフレームのピッチ成分については子音でないピッチ利得が大き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_0 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。

30

【0080】

なお、式(10),(12)においては $B_0 > B_{-\alpha} > B_{-\beta}$ とするのが好ましいが、式(10),(12)において B_0 、 $B_{-\alpha}$ や B_0 、 $B_{-\beta}$ や $B_{-\alpha}$ 、 $B_{-\beta}$ 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。

【0081】

また、式(11)と式(13)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_0 と個過去のフレームのピッチ周期 T_{-1} と個過去のフレームのピッチ周期 T_{-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。

40

【0082】

(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(5)や式(7)や式(9)や式(11)や式(11)や式(13)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_n^{new} を得るようにしてもよい。

【0083】

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。

50

【 0 0 8 4 】

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_1 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。

【 0 0 8 5 】

< 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、およびそ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、およびそ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、およびそ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を得る頻度に依存せずピッチ周期やピッチ利得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。

【 0 0 8 6 】

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。

【 0 0 8 7 】

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。

【 0 0 8 8 】

< 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。

【 0 0 8 9 】

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。

【 0 0 9 0 】

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。

【0091】

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。

10

【0092】

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。

20

【図1】

【図2】

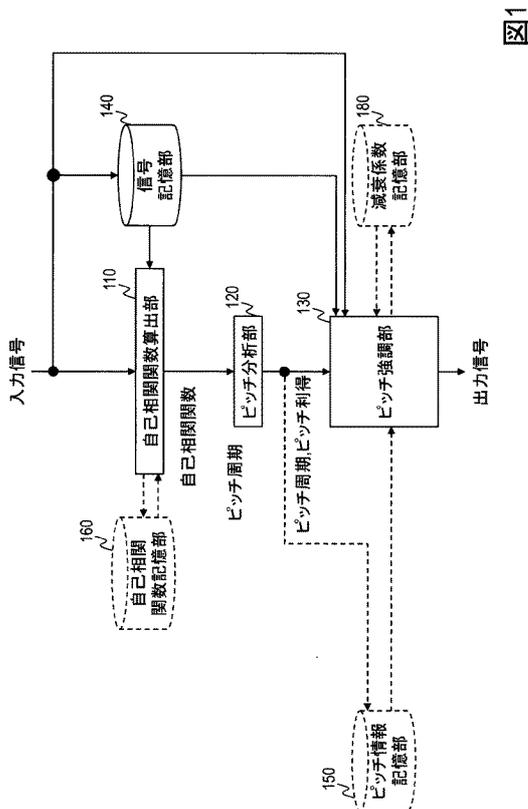


図1

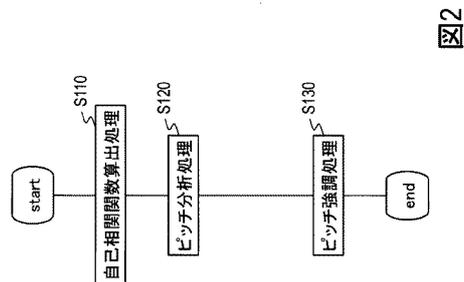


図2

【 図 3 】

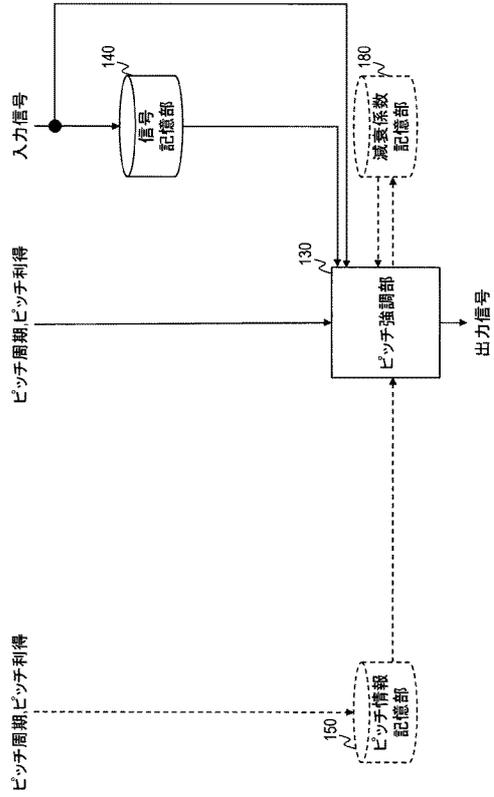


図3

【 図 4 】

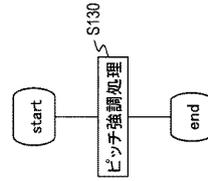


図4

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