

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第6911939号
(P6911939)

(45) 発行日 令和3年7月28日(2021.7.28)

(24) 登録日 令和3年7月12日(2021.7.12)

(51) Int. Cl. F I
G 1 O L 19/26 (2013.01) G I O L 19/26 B

請求項の数 16 (全 18 頁)

(21) 出願番号	特願2019-557077 (P2019-557077)	(73) 特許権者	000004226
(86) (22) 出願日	平成30年10月29日(2018.10.29)		日本電信電話株式会社
(86) 国際出願番号	PCT/JP2018/040150		東京都千代田区大手町一丁目5番1号
(87) 国際公開番号	W02019/107041	(74) 代理人	100121706
(87) 国際公開日	令和1年6月6日(2019.6.6)		弁理士 中尾 直樹
審査請求日	令和2年5月21日(2020.5.21)	(74) 代理人	100128705
(31) 優先権主張番号	特願2017-231747 (P2017-231747)		弁理士 中村 幸雄
(32) 優先日	平成29年12月1日(2017.12.1)	(74) 代理人	100147773
(33) 優先権主張国・地域又は機関	日本国(JP)		弁理士 義村 宗洋
		(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
		(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

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) 【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、

各時間区間について、

当該時間区間の各時刻の出力信号として、

当該時刻の前記信号と、

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、当該時刻よりも過去の時刻の、前記信号である第1強調用信号と、

当該時間区間よりも 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_1 だけ、当該時刻よりも過去の時刻の、前記信号である第2強調用信号と、

を含む信号を得る処理を行う、

ピッチ強調部を含む、

ピッチ強調装置。

【請求項2】

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の出力信号として、

当該時刻の前記信号と、

前記第 1 強調用信号に、当該時間区間のピッチ利得 σ_0 と、1 より小さい値である所定の係数 B_0 と、を乗算した信号と、

前記第 2 強調用信号に、当該時間区間よりも n 個過去の時間区間のピッチ利得 σ_{-n} と、1 より小さい値である所定の係数 B_{-n} と、を乗算した信号と、を加算した信号を含む出力信号を得るものである、
ピッチ強調装置。

【請求項 3】

請求項 2 に記載のピッチ強調装置であって、
前記係数 B_{-n} は前記係数 B_0 より小さい値である、
ピッチ強調装置。

10

【請求項 4】

請求項 2 または 3 に記載のピッチ強調装置であって、
前記ピッチ強調部は、
前記加算した信号を

【数 17】

$$\sqrt{1 + B_0^2 \sigma_0^2 + B_{-n}^2 \sigma_{-n}^2 + 2B_0 B_{-n} \sigma_0 \sigma_{-n}}$$

で除算した信号を出力信号として得るものである、
ピッチ強調装置。

20

【請求項 5】

請求項 1 のピッチ強調装置であって、
前記ピッチ強調部は、
当該時間区間のピッチ利得 σ_0 が所定の閾値より小さい場合には、前記第 1 強調用信号を出力信号に含めない、
ピッチ強調装置。

【請求項 6】

請求項 1 または請求項 5 のピッチ強調装置であって、
前記ピッチ強調部は、
当該時間区間よりも n 個過去の時間区間のピッチ利得 σ_{-n} が所定の閾値より小さい場合には、前記第 2 強調用信号を出力信号に含めない、
ピッチ強調装置。

30

【請求項 7】

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の出力信号として、
当該時間区間よりも前記 n と異なる 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_m だけ、当該時刻よりも過去の時刻の、前記信号である第 3 強調用信号も更に含む信号を得る、
ピッチ強調装置。

40

【請求項 8】

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、
各時間区間について、
当該時間区間の各時刻の出力信号として、
当該時刻の前記信号と、
当該時間区間のピッチ周期に対応するサンプル数 T_1 だけ、当該時刻よりも過去の時刻の、前記信号である第 1 強調用信号と、

50

当該時間区間よりも 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_{-} だけ、
 当該時刻よりも過去の時刻の、前記信号である第 2 強調用信号と、
 を含む信号を得る処理を行う、
 ピッチ強調ステップを含む、
 ピッチ強調方法。

【請求項 9】

請求項 8 に記載のピッチ強調方法であって、
 前記ピッチ強調ステップは、
 当該時間区間の各時刻の出力信号として、
 当該時刻の前記信号と、
 前記第 1 強調用信号に、当該時間区間のピッチ利得 G_1 と、1 より小さい値である所定
 の係数 B_1 と、を乗算した信号と、
 前記第 2 強調用信号に、当該時間区間よりも 個過去の時間区間のピッチ利得 G_2 と
 、1 より小さい値である所定の係数 B_2 と、を乗算した信号と、
 を加算した信号を含む出力信号を得る、
 ピッチ強調方法。

10

【請求項 10】

請求項 9 に記載のピッチ強調方法であって、
 前記係数 B_2 は前記係数 B_1 より小さい値である、
 ピッチ強調方法。

20

【請求項 11】

請求項 9 または 10 に記載のピッチ強調方法であって、
 前記ピッチ強調ステップは、
 前記加算した信号を

【数 18】

$$\sqrt{1 + B_0^2 \sigma_0^2 + B_{-\alpha}^2 \sigma_{-\alpha}^2 + 2B_0 B_{-\alpha} \sigma_0 \sigma_{-\alpha}}$$

で除算した信号を出力信号として得る、
 ピッチ強調方法。

30

【請求項 12】

請求項 8 のピッチ強調方法であって、
 前記ピッチ強調ステップは、
 当該時間区間のピッチ利得 G_1 が所定の閾値より小さい場合には、前記第 1 強調用信号
 を出力信号に含めない、
 ピッチ強調方法。

【請求項 13】

請求項 8 または請求項 12 のピッチ強調方法であって、
 前記ピッチ強調ステップは、
 当該時間区間よりも 個過去の時間区間のピッチ利得 G_2 が所定の閾値より小さい場
 合には、前記第 2 強調用信号を出力信号に含めない、
 ピッチ強調方法。

40

【請求項 14】

請求項 8 に記載のピッチ強調方法であって、
 前記ピッチ強調ステップは、
 当該時間区間の各時刻の出力信号として、
 当該時間区間よりも前記 T_{-} とは異なる 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_{-} だけ、当該時刻よりも過去の時刻の、前記信号である第 3 強調用信号も更に
 含む信号を得る、

50

ピッチ強調方法。

【請求項 15】

請求項 8 から請求項 14 の何れかのピッチ強調方法をコンピュータに実行させるためのプログラム。

【請求項 16】

請求項 8 から請求項 14 の何れかのピッチ強調方法をコンピュータに実行させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体。

【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、つまりピッチ成分を含むことに着目し、復号し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。

【0003】

例えば非特許文献 1 に記載されているように、ピッチ成分を強調する処理として、ピッチ周期を得る処理とそのピッチ周期を用いてピッチ成分を強調する処理とを一定の時間区間毎に行うことが一般的である。

【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

【発明の概要】

【発明が解決しようとする課題】

【0005】

しかしながら、これらの処理を行う時間区間が長い場合には、時間区間の時間長の分だけ処理遅延が増えてしまうという課題や、入力された音信号のピッチ周期が変動しているにも関わらずその変動に対応したピッチ成分を強調できないという課題がある。

【0006】

これらの課題は、上述した処理を行う時間区間を短くすれば解決できる。そして、例えば非特許文献 1 のように線形予測を伴う復号処理とピッチを強調する処理（ピッチ強調処理）とを併用する構成においては、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ことで、時間区間を短くしても更なる課題が生じることもない。しかし、線形予測を伴わない復号処理とピッチ強調処理とを併用する構成においては、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非特許文献 1 と同様の構成とするためには、新たに線形予測分析を行うために演算処理量が大幅に増加してしまうという新たな課題が生じてしまう。この新たな課題を生じさせないためには、線形予測を伴わない復号処理とピッチ強調処理とを併用する構成においては、復号処理により得られた音信号そのものに対して参考文献 1 に記載されているようなピッチ強調処理を施す必要がある。

（参考文献 1）D. Malah and R. Cox, "A generalized comb filtering technique for speech enhancement", in ICASSP 1982, vol. 7, pp. 160 163, 1982.

10

20

30

40

50

【 0 0 0 7 】

しかし、復号処理により得られた音信号そのものに対してピッチ強調処理を行う時間区間を短くしてしまうと、時間区間ごとに異なるピッチ周期でのピッチ強調処理が行われ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、という別の課題が生じてしまう。

【 0 0 0 8 】

本発明は、この課題を解決するためのものであり、復号処理により得られた音信号そのものに対して短い時間区間ごとにピッチ強調処理を施す場合であっても、音信号のピッチ周期の変動に対応するピッチ周期に基づく、音信号の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。

10

【課題を解決するための手段】

【 0 0 0 9 】

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、各時間区間について、当該時間区間の各時刻の出力信号として、当該時刻の信号と、当該時間区間のピッチ周期に対応するサンプル数 T_1 だけ、当該時刻よりも過去の時刻の、信号である第1強調用信号と、当該時間区間よりも n 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_2 だけ、当該時刻よりも過去の時刻の、信号である第2強調用信号と、を含む信号を得る処理を行う、ピッチ強調部を含む。

20

【 0 0 1 0 】

上記の課題を解決するために、本発明の他の一態様によれば、ピッチ強調方法は、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。ピッチ強調方法は、ピッチ強調処理として、各時間区間について、当該時間区間の各時刻の出力信号として、当該時刻の信号と、当該時間区間のピッチ周期に対応するサンプル数 T_1 だけ、当該時刻よりも過去の時刻の、信号である第1強調用信号と、当該時間区間よりも n 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_2 だけ、当該時刻よりも過去の時刻の、信号である第2強調用信号と、を含む信号を得る処理を行う、ピッチ強調ステップを含む。

30

【発明の効果】

【 0 0 1 1 】

本発明によれば、復号処理により得られた音信号そのものに対して短い時間区間ごとにピッチ強調処理を施す場合であっても、音信号のピッチ周期の変動に対応するピッチ周期に基づく、音信号の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】

【 0 0 1 2 】

【図1】第一実施形態に係るピッチ強調装置の機能ブロック図。

【図2】第一実施形態に係るピッチ強調装置の処理フローの例を示す図。

【図3】第一実施形態の変形例に係るピッチ強調装置の機能ブロック図。

40

【図4】第一実施形態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】

【 0 0 1 3 】

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、特に断りが無い限り、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。

【 0 0 1 4 】

< 第一実施形態のポイント >

本実施形態では、現在の時間区間の入力された音信号そのものに対してピッチ強調処理

50

を施す際に、現在の時間区間のピッチ周期に対応するピッチ成分だけを強調するのではなく、過去の時間区間のピッチ周期に対応するピッチ成分も強調する。

【 0 0 1 5 】

その際、過去の時間区間のピッチ周期に対応するピッチ成分の強調の度合いは、現在の時間区間のピッチ周期に対応するピッチ成分の強調の度合いより小さくすることが望ましい。また、過去の複数の時間区間のピッチ周期に対応するピッチ成分を強調する場合には、現在の時間区間から遠い時間区間ほど強調の度合いが小さくなるようにするとよい。ただし、それぞれの時間区間のピッチ周期に対応する成分の強調は、ピッチ利得に基づいて行う。

【 0 0 1 6 】

また、強調後のエネルギーの整合性を保つため、ピッチ強調の程度は、使用した全てのピッチ成分に基づいて決めるとよい。過去の複数の時間区間のピッチ周期に対応するピッチ成分を強調させる場合には、その複数のピッチ周期が近い値であることも想定して、強調後のエネルギーの整合性を保つようにするとよい。

【 0 0 1 7 】

< 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。

【 0 0 1 8 】

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。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 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 とピッチ情報記憶部 1 5 0 とを備えるものであり、更に自己相関関数記憶部 1 6 0 を備えてもよい。

【 0 0 1 9 】

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 (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。

【 0 0 2 0 】

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) とピッチ強調処理 (S 1 3 0) である (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。

【 0 0 2 1 】

[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 2 2 】

10

20

30

40

50

自己相関関数算出部 110 には、時間領域の音信号が入力される。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）、 \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。

【0023】

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 を備える。信号記憶部 140 は、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、1つ前のフレームまでに入力されたL - N個の音信号サンプル $X_0, X_1, \dots, X_{L-N-1}$ を記憶しておく。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L - N個の音信号サンプル $X_0, X_1, \dots, X_{L-N-1}$ を読み出し、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。

【0024】

L = Nである場合には、入力されたN個の時間領域の音信号サンプルをそのまま最新のL個の音信号サンプル X_0, X_1, \dots, X_{L-1} とすればよい。この場合、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。

【0025】

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）、 \dots 、(M)それぞれに対する自己相関関数 $R_{(1)}$, \dots , $R_{(M)}$ を算出する。（1）、 \dots 、(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_τ を例えば以下の式(1)で算出する。

【0026】

【数1】

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

【0027】

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。

【0028】

なお、この時間差（1）、 \dots 、(M)は後述するピッチ分析部 120 が求める現在のフレー

10

20

30

40

50

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 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、ピッチ周期 T_0 の候補である75から320に対して、 L を8192などの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。

【0029】

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)で求まる値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 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(直前のフレームの自己相関関数) $R^{(1)}$, ..., $R^{(M)}$ 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(直前のフレームの自己相関関数) $R^{(1)}$, ..., $R^{(M)}$ 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 τ 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 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。

【0030】

【数2】

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

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

【0031】

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。

【0032】

なお、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 $L - N$ 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 140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 140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

10

20

40

50

て新たに記憶する。

【 0 0 3 3 】

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 $R_{(1)}$, $R_{(M)}$ 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 $R_{(1)}$, $R_{(M)}$ 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。

【 0 0 3 4 】

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 $R_{(1)}, \dots, R_{(M)}$ を算出すればよく、 $L = N$ の場合と同様に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。

【 0 0 3 5 】

[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 6 】

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。

【 0 0 3 7 】

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 G 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、得たピッチ利得 G とピッチ周期 T_0 とをピッチ強調部 1 3 0 へ出力する。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 (s 個過去) のフレームのピッチ利得及びピッチ周期をそれぞれ T_{-s} 及び G_{-s} と表記する。

【 0 0 3 8 】

[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とこれに関連する処理について説明する。

【 0 0 3 9 】

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。例えば、ピッチ強調部 1 3 0 は、N サンプルにより構成される当該時間区間の信号 (現在のフレームの時間領域の音信号) $X_n (L - N \leq n < L - 1)$ に含まれる各時刻 n の信号に対応する出力信号 X_n^{out} として、時刻 n の信号 X_n と、現在のフレームのピッチ周期 T_0 に対応するサンプル数 T_0 だけ、時刻 n よりも過去の時刻である時刻 $n - T_0$ の信号 (以下、第 1 強調用信号ともいい、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である) $X_{n - T_0}$ (ただし、下付き添え字における A, B は A_0 を意味する) と、直前のフレームのピッチ周期 T_{-1} に対応するサンプル数 T_{-1} だけ、時刻 n よりも過去の時刻である時刻 $n - T_{-1}$ の信号 (以下、第 2 強調用信号ともいい、直前のフレームのピッチ周期に対応するピッチ成分である) $X_{n - T_{-1}}$ と、を含む信号を得、出力する。

【 0 0 4 0 】

以下、具体例を説明する。

【 0 0 4 1 】

10

20

30

40

50

ピッチ情報記憶部 150 には、1つ前のフレームのピッチ周期 T_{-1} とピッチ利得 σ_{-1} とを記憶しておく。

【0042】

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 σ_0 と、ピッチ情報記憶部 150 から読み出した1つ前のフレームのピッチ利得 σ_{-1} 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1つ前のフレームのピッチ周期 T_{-1} 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 ($L - N \leq n < L - 1$) 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_{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。

10

【0043】

【数3】

$$X_n^{new} = \frac{1}{A} \left[X_n + \frac{3}{4} \sigma_0 X_{n-T_0} + \frac{1}{4} \sigma_{-1} X_{n-T_{-1}} \right] \quad \dots(4)$$

【0044】

ただし、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。

【0045】

【数4】

$$A = \sqrt{1 + \frac{9}{16} \sigma_0^2 + \frac{1}{16} \sigma_{-1}^2 + \frac{3}{8} \sigma_0 \sigma_{-1}} \quad \dots(5)$$

【0046】

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る。さらに、式(4)のピッチ強調処理は、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直前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。こ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を得ることができる。

【0047】

30

具体的には、式(4)では、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 X_{n-T_0} より強調の度合いを落として直前フレームでのピッチ周期 T_{-1} に対応するピッチ成分 $X_{n-T_{-1}}$ も強調することを、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乗算する値である3/4よりも小さな値である1/4を直前フレームのピッチ周期 T_{-1} に対応するピッチ成分に乗算することで実現している。

【0048】

(変形例1)

ただし、3/4や1/4という値は一例であり、式(4)に代えて、予め定めた値である B_0 と B_{-1} を用いて下記の式(4A)により出力信号 X_n^{new} を得てもよい。

【0049】

40

【数5】

$$X_n^{new} = \frac{1}{A} \left[X_n + B_0 \sigma_0 X_{n-T_0} + B_{-1} \sigma_{-1} X_{n-T_{-1}} \right] \dots(4A)$$

【0050】

なお、式(4A)においては $B_0 > B_{-1}$ とするのが好ましいが、 $B_0 = B_{-1}$ 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。

【0051】

上記の式(5)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 と直前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。すなわち、式(4A)により出力

50

信号 X_n^{new} を得るのであれば、振幅補正係数 A として下記の式(5A)により求まる値を用いればよい。

【 0 0 5 2 】

【数 6】

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

【 0 0 5 3 】

(変形例 2)

なお、振幅補正係数 A は、式(5)や式(5A)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し、振幅補正係数 A を用いずに式(4B)や式(4C)により出力信号 X_n^{new} を得るようにしてもよい。

10

【 0 0 5 4 】

【数 7】

$$X_n^{new} = \left[X_n + \frac{3}{4} \sigma_0 X_{n-T_0} + \frac{1}{4} \sigma_{-1} X_{n-T_{-1}} \right] \quad \dots(4B)$$

$$X_n^{new} = \left[X_n + B_0 \sigma_0 X_{n-T_0} + B_{-1} \sigma_{-1} X_{n-T_{-1}} \right] \quad \dots(4C)$$

【 0 0 5 5 】

(変形例 3)

また、より過去のフレーム(2つ以上前のフレーム)におけるピッチ周期に対応するピッチ成分も強調することにより、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更に低減できる。この場合、ピッチ周期に対応するピッチ成分を強調するために加算する値に掛ける係数は、例えば下記の式(6)のように、過去のピッチ周期であればあるほど小さく設定するのがよい。そして、振幅補正係数は例えば下記の式(6)のように、より過去のフレーム(2つ以上前のフレーム)のピッチ利得も含めて算出するのがよい。

20

【 0 0 5 6 】

【数 8】

$$X_n^{new} = \frac{1}{A} \left[X_n + \frac{3}{4} \sigma_0 X_{n-T_0} + \frac{3}{16} \sigma_{-1} X_{n-T_{-1}} + \frac{1}{16} \sigma_{-2} X_{n-T_{-2}} \right]$$

$$A = \sqrt{1 + \frac{9}{16} \sigma_0^2 + \frac{9}{256} \sigma_{-1}^2 + \frac{1}{256} \sigma_{-2}^2 + \frac{9}{32} \sigma_0 \sigma_{-1} + \frac{3}{32} \sigma_0 \sigma_{-2} + \frac{3}{128} \sigma_{-1} \sigma_{-2}} \quad \dots(6)$$

【 0 0 5 7 】

ただし、式(6)においては、 T_{-1} 及び T_{-2} はそれぞれ2フレーム前のピッチ周期及びピッチ利得である。すなわち、式(6)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を強調しつつ、1つ前のフレームでのピッチ周期 T_{-1} に対応するピッチ成分と、2つ前のフレームでのピッチ周期 T_{-2} に対応するピッチ成分も強調する処理により出力信号 X_n^{new} を得る場合の例である。もちろん、式(4)に対する式(4A)や、式(5)に対する式(5A)などと同様の変更を式(6)に対して行ってもよい。つまり、予め定めた値である B_0 と B_{-1} と B_{-2} を用いて下記の式(6A)により出力信号 X_n^{new} を得るようにしてもよい。

40

【 0 0 5 8 】

【数 9】

$$X_n^{new} = \frac{1}{A} \left[X_n + B_0 \sigma_0 X_{n-T_0} + B_{-1} \sigma_{-1} X_{n-T_{-1}} + B_{-2} \sigma_{-2} X_{n-T_{-2}} \right]$$

$$A = \sqrt{1 + B_0^2 \sigma_0^2 + B_{-1}^2 \sigma_{-1}^2 + B_{-2}^2 \sigma_{-2}^2 + E + F + G} \quad (6A)$$

$$E = 2B_0 B_{-1} \sigma_0 \sigma_{-1}$$

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

$$G = 2B_{-1} B_{-2} \sigma_{-1} \sigma_{-2}$$

【0059】

これらの場合には、ピッチ情報記憶部 150 は、2 つ前のフレームのピッチ周期 T_{-2} とピッチ利得 σ_{-2} も記憶しておく。

【0060】

また、3 つ前のフレームでのピッチ周期に対応するピッチ成分なども強調する処理を行ってもよい。もちろん、この場合には、ピッチ情報記憶部 150 は、3 つ前のフレームのピッチ周期 T_{-3} とピッチ利得 σ_{-3} も記憶しておく。

【0061】

(変形例 4)

ピッチ強調部 130 の処理における、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。例えば、対称的なローパスフィルタと等価な処理を上記の式(4)の処理と共に行うのであれば、下記の式(7)により出力信号 X_n^{new} を得るようすればよい。下記の式(7)では、対称的なローパスフィルタと等価な処理を、ピッチ強調処理において各ピッチ周期分前周辺のサンプルに基づく値を加算することにより実現している。

【0062】

【数 10】

$$X_n^{new} = \frac{1}{A} \left[X_n + \frac{3}{4} \sigma_0 C + \frac{1}{4} \sigma_{-1} D \right]$$

$$C = \left(\frac{1}{8} X_{n-T_0-1} + \frac{3}{4} X_{n-T_0} + \frac{1}{8} X_{n-T_0+1} \right)$$

$$D = \left(\frac{1}{8} X_{n-T_{-1}-1} + \frac{3}{4} X_{n-T_{-1}} + \frac{1}{8} X_{n-T_{-1}+1} \right) \quad \dots(7)$$

【0063】

(変形例 5)

なお、上記のピッチ強調処理において、ピッチ利得が所定の閾値 Th 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、式(4)のピッチ強調処理を行う際に、現在のフレームのピッチ利得 σ_0 が閾値 Th より小さい値であり、1 つ前のフレームのピッチ利得 σ_{-1} が閾値 Th 以上の値である場合には、下記の式(8)により出力信号 X_n^{new} を得るようにしてもよい。

【0064】

10

20

40

【数 1 1】

$$X_n^{new} = \frac{1}{A} \left[X_n + \frac{1}{4} \sigma_{-1} X_{n-T_{-1}} \right] \quad \left(A = \sqrt{1 + \frac{1}{16} \sigma_{-1}^2} \right) \quad \dots (8)$$

【 0 0 6 5】

また、例えば、式(6)のピッチ強調を行う際に、現在のフレームのピッチ利得 G_n が閾値 Th 以上の値であり、1つ前のフレームのピッチ利得 $G_{n-T_{-1}}$ が閾値 Th より小さい値であり、2つ前のフレームのピッチ利得 $G_{n-T_{-2}}$ が閾値 Th 以上の値である場合には、下記の式(9)により出力信号 X_n^{new} を得ればよい。

10

【 0 0 6 6】

【数 1 2】

$$X_n^{new} = \frac{1}{A} \left[X_n + \frac{3}{4} \sigma_0 X_{n-T_0} + \frac{1}{16} \sigma_{-2} X_{n-T_{-2}} \right]$$

$$A = \sqrt{1 + \frac{9}{16} \sigma_0^2 + \frac{1}{256} \sigma_{-2}^2 + \frac{3}{32} \sigma_0 \sigma_{-2}} \quad \dots (9)$$

【 0 0 6 7】

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とピッチ利得を、次のフレームのピッチ強調部 130 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。例えば、ピッチ強調部 130 が式(4)の処理を行う場合には、ピッチ情報記憶部 150 は、記憶されている1つ前のフレームのピッチ周期 T_{-1} とピッチ利得 G_{-1} を削除し、代わりに、現在のフレームのピッチ周期 T_0 をピッチ周期 T_{-1} として新たに記憶し、現在のフレームのピッチ利得 G_0 をピッチ利得 G_{-1} として新たに記憶する。同様に、ピッチ強調部 130 が式(6)の処理を行う場合には、ピッチ情報記憶部 150 は、記憶されている2つ前のフレームのピッチ周期 T_{-2} とピッチ利得 G_{-2} を削除し、記憶されている1つ前のフレームのピッチ周期 T_{-1} とピッチ利得 G_{-1} のそれぞれを T_{-2} と G_{-2} とし、現在のフレームのピッチ周期 T_0 とピッチ利得 G_0 のそれぞれを T_{-1} と G_{-1} として新たに記憶する。

20

30

【 0 0 6 8】

(変形例 6)

上述の第2強調用信号は必ずしも直前のフレームのピッチ周期に対応するピッチ成分でなくともよく、現在のフレームよりも M 個過去のフレームのピッチ周期に対応するピッチ成分であればよい。ただし、 M は1以上の整数の何れかである。つまり、ピッチ強調部 130 は、現在のフレームの時間領域の音信号サンプル X_n ($L - N \leq n \leq L - 1$) に含まれる各時刻 n の信号に対応する出力信号 X_n^{new} として、時刻 n の信号 X_n と、現在のフレームのピッチ周期 T_0 に対応するサンプル数 T_0 だけ、時刻 n よりも過去の時刻である時刻 $n - T_0$ の信号 (以下、第1強調用信号ともいい、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である) X_{n-T_0} と、現在のフレームよりも M 個過去のフレームのピッチ周期 T_{-M} に対応するサンプル数 T_{-M} だけ、時刻 n よりも過去の時刻である時刻 $n - T_{-M}$ の信号 (以下、第2強調用信号ともいい、 M 個過去のフレームのピッチ周期に対応するピッチ成分である) $X_{n-T_{-M}}$ と、を含む信号を得てもよい。例えば、現在のフレームの音信号サンプル X_n ($L - N \leq n \leq L - 1$) に含まれる各時刻 n の信号に対応する出力信号 X_n^{new} として、時刻 n の信号 X_n と、第1強調用信号 X_{n-T_0} に、現在のフレームのピッチ利得 G_0 と、1より小さい値である所定の係数 B_0 と、を乗算した信号と、第2強調用信号 $X_{n-T_{-M}}$ に、現在のフレームよりも M 個過去のフレームのピッチ利得 G_{-M} と、1より小さい値である所定の係数 B_{-M} と、を乗算した信号と、を加算した信号を

40

【 0 0 6 9】

50

【数 1 3】

$$A = \sqrt{1 + B_0^2 \sigma_0^2 + B_{-\alpha}^2 \sigma_{-\alpha}^2 + 2B_0 B_{-\alpha} \sigma_0 \sigma_{-\alpha}}$$

【0 0 7 0】

で除算した信号を得る処理を行ってもよい。つまり、

【0 0 7 1】

【数 1 4】

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

【0 0 7 2】

である。変形例 1 と同様に $B_0 > B_{-\alpha}$ とするのが好ましいが、 $B_0 = B_{-\alpha}$ 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。さらに、この変形例と上述の変形例 2 ~ 5 とを組合せてもよい。

【0 0 7 3】

つまり、変形例 2 と組合せて、振幅補正係数 A を予め定めた 1 以上の値としてもよいし、振幅補正係数 A を用いなくともよい。

【0 0 7 4】

変形例 3 と組合せて、現在のフレームの時間領域の音信号サンプル X_i ($L - N - n - L - 1$) に含まれる各時刻 n の信号に対応する出力信号 X_n^{new} として、時刻 n の信号 X_n と、第 1 強調用信号 X_{n-T_0} と、第 2 強調用信号 $X_{n-T_{-\alpha}}$ とに加え、現在のフレームよりも 1 個過去のフレームのピッチ周期 T_{-1} に対応するサンプル数 T_{-1} だけ、時刻 n よりも過去の時刻である時刻 $n - T_{-1}$ の信号 (以下、第 3 強調用信号ともいい、1 個過去のフレームのピッチ周期に対応するピッチ成分である) $X_{n-T_{-1}}$ も更に含む信号を得てもよい。ただし、 $T_{-1} > T_0$ である。例えば、 $T_{-1} = 2T_0$ とすると、

【0 0 7 5】

【数 1 5】

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

$$A = \sqrt{1 + B_0^2 \sigma_0^2 + B_{-\alpha}^2 \sigma_{-\alpha}^2 + B_{-\beta}^2 \sigma_{-\beta}^2 + E + F + G}$$

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

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

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

【0 0 7 6】

である。 $B_0 > B_{-\alpha} > B_{-\beta}$ 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。これらの場合には、ピッチ情報記憶部 150 は、1 個前のフレームまでのピッチ周期 $T_0, T_{-1}, \dots, T_{-Q}$ とピッチ利得 $\sigma_0, \sigma_{-1}, \dots, \sigma_{-Q}$ を記憶しておく。また、現在のフレームの時間領域の音信号サンプル X_i ($L - N - n - L - 1$) に含まれる各時刻 n の信号に対応する出力信号 X_n^{new} として、時刻 n の信号 X_n と、第 1 強調用信号 X_{n-T_0} と、第 2 強調用信号 $X_{n-T_{-\alpha}}$ と、第 3 強調用信号 $X_{n-T_{-\beta}}$ とに加え、Q 個の第 (q + 3) 強調用信号

【0 0 7 7】

【数 1 6】

$$X_{n-T_{-q}}$$

【0 0 7 8】

を更に含む信号を得てもよい。ただし、 $q=1, 2, \dots, Q$ であり、Q は 1 以上の整数の何れかであり、 $T_{-q} > T_{-(q-1)} > T_{-(q-2)} > \dots > T_{-1}$ とする。もちろん、この場合には、ピッチ情報記憶部 150 は

10

20

40

50

、最も古い $(n-1)$ 個前のフレームまでのピッチ周期 $T_0, T_{-1}, \dots, T_{-(n-1)}$ とピッチ利得 $G_0, G_{-1}, \dots, G_{-(n-1)}$ を記憶しておく。

【0079】

変形例4と組合せて、ピッチ強調部130の処理における、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。

【0080】

変形例5と組合せて、現在のフレームのピッチ利得 G_n が所定の閾値より小さい場合には、第1強調用信号を出力信号に含めず、現在のフレームよりも $(n-1)$ 個過去のフレームのピッチ利得 G_{n-1} が所定の閾値より小さい場合には、第2強調用信号を出力信号に含めない構成としてもよい。

【0081】

さらに、各変形例を組合せてもよい。例えば、変形例3と変形例5と変形例6とを組合せて、現在のフレームよりも $(n-2)$ 個過去のフレームのピッチ利得 G_{n-2} が所定の閾値より小さい場合には、第3強調用信号を出力信号に含めない構成としてもよい。

【0082】

<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、音声ピッチ強調装置は、第一実施形態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を備える必要はない。この場合、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に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。

【0083】

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。

【0084】

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。

【0085】

<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 6 】

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 8 7 】

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。

10

【 0 0 8 8 】

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。

20

【 0 0 8 9 】

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。

【図1】

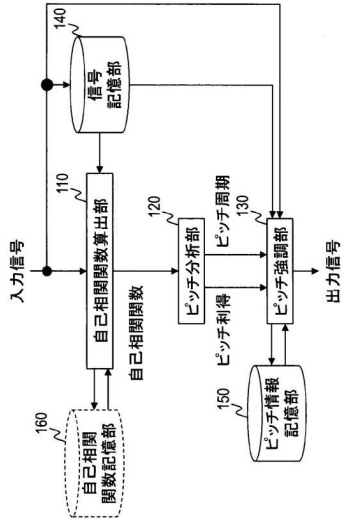


図1

【図2】

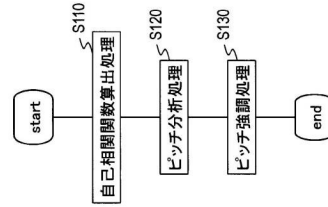


図2

【図3】

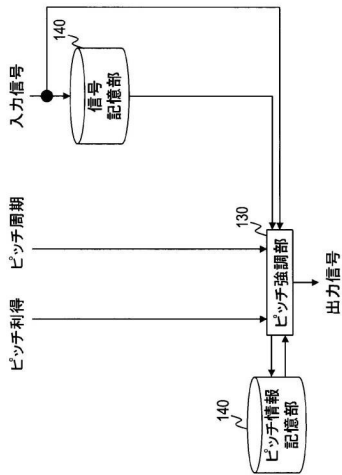


図3

【図4】

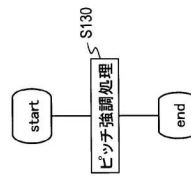


図4

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

審査官 菊池 智紀

(56)参考文献 特開平11-272297(JP,A)

特開2001-147700(JP,A)

特開2000-148194(JP,A)

千葉大将 他, "CELPに基づく音声符号化向けのピッチ周波数に依存した適応ポストフィルタ",
電子情報通信学会論文誌D, 2015年10月01日, Vol.J98 D, No.10, pp.1301-1311

(58)調査した分野(Int.Cl., DB名)

G10L 19/00 - 21/0364

IEEE Xplore