

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第6616962号  
(P6616962)

(45) 発行日 令和1年12月4日(2019.12.4)

(24) 登録日 令和1年11月15日(2019.11.15)

(51) Int.Cl.		F I	
<b>G 1 O L</b>	<b>21/007</b>	<b>(2013.01)</b>	G 1 O L 21/007
<b>G 1 O L</b>	<b>25/12</b>	<b>(2013.01)</b>	G 1 O L 25/12
<b>G 1 O L</b>	<b>25/15</b>	<b>(2013.01)</b>	G 1 O L 25/15
<b>G 1 O L</b>	<b>25/18</b>	<b>(2013.01)</b>	G 1 O L 25/18

請求項の数 2 (全 18 頁)

(21) 出願番号	特願2015-98608 (P2015-98608)	(73) 特許権者	000004352 日本放送協会 東京都渋谷区神南2丁目2番1号
(22) 出願日	平成27年5月13日(2015.5.13)	(73) 特許権者	591053926 一般財団法人NHKエンジニアリングシステム 東京都世田谷区砧一丁目10番11号
(65) 公開番号	特開2016-212356 (P2016-212356A)	(74) 代理人	110001564 フェリシテ特許業務法人
(43) 公開日	平成28年12月15日(2016.12.15)	(72) 発明者	清山 信正 東京都世田谷区砧一丁目10番11号 日本放送協会放送技術研究所内
審査請求日	平成30年4月2日(2018.4.2)	(72) 発明者	今井 篤 東京都世田谷区砧一丁目10番11号 日本放送協会放送技術研究所内

最終頁に続く

(54) 【発明の名称】 信号処理装置及びプログラム

(57) 【特許請求の範囲】

【請求項1】

入力された音声信号に対して声質変換処理を施すように構成されている信号処理装置であって、

前記音声信号から周波数特性を算出する周波数特性算出部と、

前記音声信号から算出された線形予測係数に基づいて、前記音声信号のスペクトル包絡を算出するスペクトル包絡算出部と、

周波数変換を行う周波数範囲を指定する周波数変換パラメータに基づいて、前記周波数範囲内の入力周波数を周波数変換するための周波数変換関数を作成する周波数変換関数作成部と、

前記周波数変換関数を用いて、前記スペクトル包絡を構成する各周波数成分を前記入力周波数として前記周波数変換をすることにより前記スペクトル包絡を変更する変更部と、

前記周波数成分が変更された前記スペクトル包絡に基づいて、前記周波数特性を補正する補正部と、

補正された前記周波数特性に基づいて、前記声質変換処理が施された音声信号を取得する処理部と

を具備することを特徴とする信号処理装置。

【請求項2】

前記スペクトル包絡算出部は、前記線形予測係数として、前記入力された音声信号をダウンサンプリングすることによって得られた音声帯域部分のみを対象とした線形予測係数

を用いることを特徴とする請求項 1 に記載の信号処理装置。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、入力された音声信号に対して声質変換処理を施す信号処理装置及びプログラムに関する。

【背景技術】

【0002】

従来、音声信号に対して所定処理を施すことによって声質を変換する方法として、様々な方法が提案されている。例えば、音声信号からスペクトル包絡を抽出して声質を変換する方法として、以下に示す特許文献 1 に記載されている方法が知られている。

【0003】

かかる特許文献 1 に記載されている方法は、以下の手順によって行われる。第 1 に、入力された音声信号から有声音区間を抽出する。第 2 に、抽出された有声音区間において、線形予測係数を求めることによって、ホルマント周波数（共振周波数）及び帯域幅を算出すると共に、線形予測係数に基づいてスペクトル包絡（変更前スペクトル包絡）を算出する。第 3 に、有声音区間における音声信号に対してフーリエ変換を施して周波数領域の成分（周波数成分）に変換する。第 4 に、ホルマント周波数の各時間軌跡におけるホルマント周波数又は帯域幅に変更を加える。第 5 に、変更が加えられたホルマント周波数又は帯域幅に基づいてスペクトル包絡（変更後スペクトル包絡）を算出する。第 6 に、変更後スペクトル包絡を変更前スペクトル包絡で除した商（変更成分）を算出する。第 7 に、上述の周波数成分に対して、かかる変更成分を乗ずることにより、上述のホルマント周波数に対する変更を含むスペクトル変更を行った後、かかる周波数成分に対して逆フーリエ変換を施して時間領域の成分の音声信号を取得する。第 8 に、かかる音声信号を、無声音区間、無音区間又は前後の有声音区間に接続して、声質変更が施された音声信号を取得する。

【先行技術文献】

【特許文献】

【0004】

【特許文献 1】特許第 2 6 1 2 8 6 9 号

【発明の概要】

【発明が解決しようとする課題】

【0005】

ここで、上述の特許文献 1 に記載されている方法では、変換ビット数が 12 ビットで標本化周波数が 15 kHz で A/D 変換された音声信号を対象とした実施例が挙げられており、かかる実施例では、男性の音声信号（男声）に対しては線形予測分析次数  $p$  を 14 次程度とし、女性の音声信号（女性）に対しては線形予測分析次数  $p$  を 10 次程度としている。

【0006】

図 11 に、特許文献 1 に記載されている方法で用いられているパラメータとほぼ同じパラメータとして標本化周波数を 16 kHz とし線形予測分析次数を 16 次とした場合のスペクトル包絡を示す。ここで、図 11 では、横軸が周波数（Hz）を示し、縦軸が振幅（dB）を示し、実線がパワースペクトルを示し、点線がスペクトル包絡を示す。

【0007】

ここで、線形予測分析によるスペクトル包絡が、パワースペクトルのピーク部分を捉えており、声質を変換する際に、個人性に寄与する主に 4 kHz までの低い周波数から 3 個程度のピークを変化させることができる。

【0008】

図 12 に、標本化周波数を 48 kHz とし線形予測分析次数を 16 次とした場合のスペクトル包絡を示す。

【0009】

10

20

30

40

50

かかる場合には、すなわち、線形予測分析次数として16次を用いて標本化周波数48kHzという広い帯域の音声信号に対して線形予測分析を行った場合には、細かい共振のピークを求めることができないため、標本化周波数が15kHzとされているケースで想定しているような音声帯域のピーク部分を捉えることができない。

【0010】

一方で、図13に、標本化周波数を48kHzとして線形予測分析次数を48次とした場合のスペクトル包絡を示す。

【0011】

かかる場合には、すなわち、線形予測分析次数を増やして適切な次数で標本化周波数が48kHzといった広い帯域の音声信号に対して線形予測分析を行った場合には、標本化周波数が15kHzとされているケースで想定しているようなホルマント周波数に相当する音声帯域も含めて、細かい共振のピークを求めることができる。

10

【0012】

以上のように、上述の特許文献1に記載されている方法では、帯域幅が15kHz以上の広い帯域の音声信号（例えば、標本化周波数が48kHzの音声信号）を対象として、標本化周波数が15kHzとされているケースで想定しているようなホルマント周波数に相当する音声帯域のスペクトル包絡を変更した音声を得ることができないという問題点があった。

【0013】

そこで、本発明は、上述した課題を解決するためになされたものであり、入力された音声信号に対して高品質な声質変換を行うことができる信号処理装置及びプログラムを提供することを目的とする。

20

【課題を解決するための手段】

【0014】

本発明の第1の特徴は、入力された音声信号に対して所定処理を施すように構成されている信号処理装置であって、前記音声信号から周波数特性を算出するように構成されている周波数特性算出部と、前記音声信号から算出された線形予測係数に基づいて、前記音声信号のスペクトル包絡を算出するように構成されているスペクトル包絡算出部と、入力された周波数を一意に変換する周波数変換関数を用いて、前記スペクトル包絡を構成する周波数成分を変更するように構成されている変更部と、前記周波数成分が変更された前記スペクトル包絡に基づいて、前記周波数特性を補正するように構成されている補正部と、補正された前記周波数特性に基づいて、前記所定処理が施された音声信号を取得するように構成されている処理部とを具備することを要旨とする。

30

【0015】

本発明の第2の特徴は、入力された音声信号に対して所定処理を施すように構成されている信号処理装置であって、前記音声信号から周波数特性を算出するように構成されている周波数特性算出部と、前記音声信号から算出された線形予測係数に基づいて、前記音声信号のスペクトル包絡を算出するように構成されているスペクトル包絡算出部と、ホルマントに所望の変更を加えることによって、前記スペクトル包絡に対して所望の変更を加えるように構成されている変更部と、前記所望の変更が加えられた前記スペクトル包絡に基づいて、前記周波数特性を補正するように構成されている補正部と、補正された前記周波数特性に基づいて、前記所定処理が施された音声信号を取得するように構成されている処理部とを具備し、前記スペクトル包絡算出部は、前記線形予測係数として、前記入力された音声信号をダウンサンプリングすることによって得られた音声帯域部分のみを対象とした線形予測係数を用いるように構成されていることを要旨とする。

40

【0016】

本発明の第3の特徴は、コンピュータを、上述の信号処理装置として機能させるためのプログラムであることを要旨とする。

【発明の効果】

【0017】

50

本発明によれば、入力された音声信号に対して高品質な声質変換を行うことができる信号処理装置及びプログラムを提供することができる。

【図面の簡単な説明】

【0018】

【図1】図1は、第1の実施形態に係る信号処理装置1の機能ブロック図である。

【図2】図2は、第1の実施形態に係る信号処理装置1で用いられる周波数変換関数の一例を示す図である。

【図3】図3(a)は、第1の実施形態に係る信号処理装置1によって声質が変換された音声信号のスペクトル包絡を示し、図3(b)は、第1の実施形態に係る信号処理装置1によって声質が変換された音声信号のパワースペクトルを示す(周波数変換倍率 = 0.5)。

10

【図4】図4(a)は、第1の実施形態に係る信号処理装置1によって声質が変換された音声信号のスペクトル包絡を示し、図4(b)は、第1の実施形態に係る信号処理装置1によって声質が変換された音声信号のパワースペクトルを示す(周波数変換倍率 = 2.0)。

【図5】図5は、第2の実施形態に係る信号処理装置1の機能ブロック図である。

【図6】図6(a)は、第2の実施形態に係る信号処理装置1によって声質が変換された音声信号の音声帯域部分におけるスペクトル包絡を示し、図6(b)は、第2の実施形態に係る信号処理装置1によって声質が変換された音声信号の音声帯域部分におけるパワースペクトルを示す(周波数変換倍率 = 0.5)。

20

【図7】図7(a)は、第2の実施形態に係る信号処理装置1によって声質が変換された音声信号の音声帯域部分におけるスペクトル包絡を示し、図7(b)は、第2の実施形態に係る信号処理装置1によって声質が変換された音声信号の音声帯域部分におけるパワースペクトルを示す(周波数変換倍率 = 2.0)。

【図8】図8は、第3の実施形態に係る信号処理装置1の機能ブロック図である。

【図9】図9(a)は、第3の実施形態に係る信号処理装置1によって声質が変換された音声信号の音声帯域部分におけるスペクトル包絡を示し、図9(b)は、第3の実施形態に係る信号処理装置1によって声質が変換された音声信号の音声帯域部分におけるパワースペクトルを示す(共振周波数に対するホルマント制御倍率 = 0.5)。

【図10】図10(a)は、第3の実施形態に係る信号処理装置1によって声質が変換された音声信号の音声帯域部分におけるスペクトル包絡を示し、図10(b)は、第3の実施形態に係る信号処理装置1によって声質が変換された音声信号の音声帯域部分におけるパワースペクトルを示す(共振周波数に対するホルマント制御倍率 = 2.0)。

30

【図11】図11は、従来技術について説明するための図である。

【図12】図12は、従来技術について説明するための図である。

【図13】図13は、従来技術について説明するための図である。

【発明を実施するための形態】

【0019】

(第1の実施形態)

以下、図1～図4を参照して、本発明の第1の実施形態に係る信号処理装置1について説明する。

40

【0020】

図1に示すように、本実施形態に係る信号処理装置1は、フレーム切出部11と、フーリエ変換部12と、線形予測分析部13と、スペクトル包絡算出部14と、スペクトル包絡変更部15と、周波数変換関数作成部16と、スペクトル包絡減算部17と、スペクトル包絡補正部18と、スペクトル包絡減算傾き算出部19と、スペクトル包絡傾き補正部20と、逆フーリエ変換部21と、フレーム合成部22とを具備している。

【0021】

フレーム切出部11は、入力された音声信号から、適切な長さのフレーム単位でデータを切り出すように構成されている。

50

## 【 0 0 2 2 】

例えば、フレーム切出部 1 1 は、標本化周波数  $f_s$  が 4 8 k H z で且つ量子化ビット数が 1 6 ビットで A / D 変換されて標本化された音声信号（音声波形）を取得すると、適切な長さのフレーム単位でデータを切り出すように構成されている。

## 【 0 0 2 3 】

なお、第 1 の実施形態の例では、入力された音声信号を  $x_i(n)$  で表し、フレーム長  $N$  を 1 0 2 4 サンプル（約 2 0 m s）とし、フレームシフト幅をフレーム長  $N$  の半分の 5 1 2 サンプル（約 1 0 m s）とし、 $m$  番目のフレームで  $n$  番目のサンプルのデータを  $x_i(m, n)$  で表す。

## 【 0 0 2 4 】

また、フレーム切出部 1 1 は、音声信号における声道特性特有の高域の減衰を補正するため、切り出したフレーム内のデータに、「 $P(z) = 1 - z^{-1}$  ( $\alpha = 0.97$ )」で表されるフィルタで、プリアンファシスを施すように構成されていてもよい。

## 【 0 0 2 5 】

また、フレーム切出部 1 1 は、プリアンファシスを施したデータに対して、適切な窓関数（例えば、ハミング窓  $w_{hamm}(n) = 0.54 - 0.46 \cos(2\pi n/N)$ ）を乗じるように構成されていてもよい。

## 【 0 0 2 6 】

フーリエ変換部 1 2 は、フレーム切出部 1 1 から、入力された音声信号から切り出されてプリアンファシスを施されて窓関数を乗じられたデータを取得すると、離散フーリエ変換（FFT）によって、周波数特性  $X(z)$  を算出するように構成されている。なお、本実施形態の例では、FFT ポイント数を 2 0 4 8 ポイントとする。

## 【 0 0 2 7 】

また、線形予測分析部 1 3 は、フレーム切出部 1 1 から、入力された音声信号から切り出されたデータを取得すると、かかるデータに基づく線形予測分析により、全ての帯域を対象とした線形予測係数  $a_i$  ( $i = 1, \dots, p$ ) を算出するように構成されている。本実施形態の例では、線形予測係数の次数（線形予測分析次数） $p$  を 4 8 次とする。

## 【 0 0 2 8 】

スペクトル包絡算出部 1 4 は、線形予測分析部 1 3 から線形予測係数を取得すると、スペクトル包絡

## 【 数 1 】

$$H(z) = \log \left[ 1 / \left( 1 + \sum_{i=1}^p a_i \cdot z^{-i} \right) \right] \quad (\text{ここでは、複素対数をとっている})$$

を算出するように構成されている。本実施形態の例では、スペクトル包絡のデータ長は、周波数特性の FFT サンプル数と同じ 2 0 4 8 ポイントである。

## 【 0 0 2 9 】

周波数変換関数作成部 1 6 は、入力された周波数変換パラメータに基づいて、周波数変換関数を作成するように構成されている。

## 【 0 0 3 0 】

ここで、図 2 に、参考文献 1（「The HTK Book (for HTK Version 3.4) Cambridge University Engineering Department」）に示されている周波数変換関数の図に、説明を加えた例を示す。図 2 では、横軸がオリジナルの入力周波数  $f_{in}$  を示し、縦軸が周波数変換された出力周波数  $f_{out}$  を示す。なお、参考文献 1 では、周波数変換関数は、音声認識における声道の長さの影響を正規化するために利用されている。

## 【 0 0 3 1 】

本実施形態では、周波数変換関数作成部 1 6 は、例えば、図 2 に示す周波数変換関数を

10

20

30

40

50

用いるように構成されていてもよい。かかる周波数変換関数は、入力された周波数  $f_{in}$  を一意に変換するための関数、すなわち、入力周波数  $f_{in}$  を入力周波数  $f_{in}$  に対応する出力周波数  $f_{out}$  に変換する関数である。

【0032】

また、かかる場合、周波数変換関数作成部16に入力される周波数変換パラメータは、境界最低周波数  $f_L$ 、境界最高周波数  $f_U$ 、変換範囲指定最低周波数  $f_{min}$ 、変換範囲指定最高周波数  $f_{max}$  及び周波数変換倍率  $= 1 / ( \text{声道長補正倍率} \cdot \frac{f_{min}}{f_{max}} )$  が含まれる。

【0033】

ここで、例えば、周波数変換関数作成部16に対して入力される周波数変換パラメータとして、 $f_L = 100$ 、 $f_U = 5000$ 、 $f_{min} = 0$ 、 $f_{max} = 24000$  と設定し、 $0.5 \sim 2.0$  の範囲で  $\text{声道長補正倍率}$  を設定する。なお、周波数制御倍率については、図2に示すように、 $\text{声道長補正倍率}$  が1より小さい程 ( $\text{声道長補正倍率}$  が1より大きい程) 入力周波数に対して出力周波数は小さくなり、 $\text{声道長補正倍率}$  が1より大きい程 ( $\text{声道長補正倍率}$  が1より小さい程) 入力周波数に対して出力周波数は大きくなる。

10

【0034】

なお、周波数変換関数作成部16に対して入力される周波数変換パラメータとして、 $f_L = 100$ 、 $f_U = 5000$ 、 $f_{min} = 0$ 、 $f_{max} = 6000$  と設定することにより、 $6000 \text{ Hz}$  以下の音声帯域部分のみを周波数変換し、 $6000 \text{ Hz}$  以上の帯域部分については、原音の周波数特性を保つこともできる。

20

【0035】

スペクトル包絡変更部15は、スペクトル包絡算出部から取得したスペクトル包絡  $H(z)$  及び周波数変換関数作成部16から取得した周波数変換関数に基づいて、スペクトル包絡  $H(z)$  を変更することによってスペクトル包絡  $H'(z)$  を算出するように構成されている。

【0036】

例えば、スペクトル包絡変更部15は、図2に示す周波数関数を用いて、スペクトル包絡  $H'(z)$  を算出するように構成されていてもよい。かかる場合、スペクトル包絡変更部15は、スペクトル包絡  $H(z)$  を構成する各周波数成分を入力周波数  $f_{in}$  として図2に示す周波数関数に入力し、図2に示す周波数関数によって得られた複数の出力周波数  $f_{out}$  を周波数成分とするスペクトル包絡  $H'(z)$  を算出するように構成されている。

30

【0037】

ここで、スペクトル包絡変更部15は、サンプル値の補完によって、スペクトル包絡  $H(z)$  を変更するように構成されていてもよい。例えば、スペクトル包絡変更部15は、スペクトル包絡  $H(z)$  を構成する周波数成分を入力周波数  $f_{in}$  として、かかる入力周波数  $f_{in}$  に対応する出力周波数  $f_{out}$ 、及び、その前後の周波数に対応する3点のスペクトル包絡の値を用いて、2次の多項式で最小二乗近似することにより、スペクトル包絡  $H'(z)$  を算出するように構成されていてもよい。

【0038】

スペクトル包絡減算部17は、スペクトル包絡算出部14から取得したスペクトル包絡  $H(z)$  及びスペクトル包絡変更部15から取得した変更後のスペクトル包絡  $H'(z)$  に基づいて、変更後のスペクトル包絡  $H'(z)$  からスペクトル包絡  $H(z)$  を減算した補正值  $G(z) = H'(z) - H(z)$  を算出するように構成されている。ここで、スペクトル包絡については対数が取られているので、上述の減算によって補正值  $G(z)$  が求められる。

40

【0039】

スペクトル包絡補正部18は、フーリエ変換部12から取得した周波数特性  $X(z)$  及びスペクトル包絡減算部17から取得した補正值  $G(z)$  に基づいて、周波数特性  $X(z)$  に対して補正值  $G(z)$  の複素指数を取ったものを乗算することにより、補正された周

50

波数特性  $X'(z) = X(z) \cdot \exp(G(z))$  を算出するように構成されている。

【0040】

スペクトル包絡減算傾き算出部 19 は、スペクトル包絡減算部 17 から取得した補正值  $G(z)$  に基づいて、補正值  $G(z)$  の傾き  $G'(z)$  を算出するように構成されている。

【0041】

例えば、スペクトル包絡減算傾き算出部 19 は、補正值  $G(z)$  を 5 次の曲線で最小二乗近似することにより、上述の傾き  $G'(z)$  を算出するように構成されていてもよい。

【0042】

スペクトル包絡減算傾き補正部 20 は、スペクトル包絡補正部 18 から取得した補正後の周波数特性  $X'(z)$  及びスペクトル包絡減算傾き算出部 19 から取得した傾き  $G'(z)$  に基づいて、補正後の周波数特性  $X'(z)$  に対して傾き  $G'(z)$  の複素指数を取ったものを乗算することにより、補正後の周波数特性の傾き  $X''(z) = X'(z) \cdot \exp(-G'(z))$  を算出するように構成されている。

10

【0043】

逆フーリエ変換部 21 は、スペクトル包絡減算傾き補正部 20 から取得された補正後の周波数特性の傾き  $X''(z)$  に対して逆フーリエ変換 (IFFT) を施すことによって、フレームごとに声質変換された音声信号の時間波形  $x'(m, n)$  を算出するように構成されている。

【0044】

フレーム合成部 22 は、逆フーリエ変換部 21 から取得した音声信号の時間波形  $x'(m, n)$  をハミング窓  $w_{hamm}(n)$  で除して、プリエンファシスの逆フィルタでディエンファシスを施すように構成されている。

20

【0045】

ここで、フレーム合成部 22 は、ディエンファシスを施したデータに対して、適切な窓関数 (例えば、ハミング窓  $w_{hamm}(n) = 0.5 - 0.5 \cos(2\pi n/N)$ ) を乗じるように構成されている。

【0046】

その後、フレーム合成部 22 は、かかる窓関数を乗じたデータをフレームごとにシフト幅分ずらして重ね合わせるにより、声質変換がなされた音声信号  $x_o(n)$  を算出するように構成されている。

30

【0047】

ここで、図 3 (a) に、周波数変換倍率  $\alpha$  を 0.5 とした場合に本実施形態に係る信号処理装置 1 によって声質が変換された音声信号のスペクトル包絡を示し、図 3 (b) に、周波数変換倍率  $\alpha$  を 0.5 とした場合に本実施形態に係る信号処理装置 1 によって声質が変換された音声信号のパワースペクトルを示す。

【0048】

一方、図 4 (a) に、周波数変換倍率  $\alpha$  を 2.0 とした場合に本実施形態に係る信号処理装置 1 によって声質が変換された音声信号のスペクトル包絡を示し、図 4 (b) に、周波数変換倍率  $\alpha$  を 2.0 とした場合に本実施形態に係る信号処理装置 1 によって声質が変換された音声信号のパワースペクトルを示す。

40

【0049】

本実施形態の特徴は、入力された音声信号  $x_i(n)$  に対して声質変換処理 (所定処理) を施すように構成されている信号処理装置 1 であって、音声信号  $x_i(n)$  から周波数特性  $X(z)$  を算出するように構成されているフーリエ変換部 (周波数特性算出部) 12 と、音声信号  $x_i(n)$  から算出された線形予測係数  $a_i$  に基づいて、音声信号  $x_i(n)$  のスペクトル包絡  $H(z)$  を算出するように構成されているスペクトル包絡算出部 14 と、入力された周波数  $f_{in}$  を一意に変換する周波数変換関数を用いて、スペクトル包絡  $H(z)$  を構成する周波数成分を変更するように構成されているスペクトル包絡変更部 (変更部) 15 と、周波数成分が変更されたスペクトル包絡  $H'(z)$  に基づいて、周波数

50

特性  $X(z)$  を補正するように構成されているスペクトル包絡補正部 18 及びスペクトル包絡傾き補正部 20 (補正部) と、補正された周波数特性  $X'(z)$  に基づいて、声質変換処理が施された音声信号  $x_o(n)$  を取得するように構成されている逆フーリエ変換部 21 及びフレーム合成部 22 (処理部) とを具備することを要旨とする。

【0050】

ここで、かかる本実施形態の特徴において、スペクトル包絡算出部 14 は、上述の線形予測係数  $a_i$  として、入力された音声信号  $x_i(n)$  の全ての帯域を対象とした線形予測係数  $a_i$  を用いるように構成されている。

【0051】

本実施形態に係る信号処理装置 1 によれば、周波数変換関数を利用することにより、どのような波形の音声信号が入力された場合であっても、高品質に声質変換を行うことができる。

10

【0052】

(第2の実施形態)

以下、図5～図8を参照して、本発明の第2の実施形態に係る信号処理装置1について、上述の第1の実施形態に係る信号処理装置1との相違点に着目して説明する。

【0053】

本実施形態に係る信号処理装置1は、入力された音声信号  $x_i(n)$  に対する声質変換処理を行うに際して、上述の第1の実施形態に係る信号処理装置1とは異なり、入力された音声信号  $x_i(n)$  をダウンサンプリングした後、ダウンサンプリングされた帯域部分 (音声帯域部分) のみを対象とした線形予測係数  $a_i$  を用いるように構成されている。

20

【0054】

図5に示すように、本実施形態に係る信号処理装置1は、フレーム切出部11と、フーリエ変換部12と、ダウンサンプリング部31と、線形予測分析部13と、スペクトル包絡算出部14と、スペクトル包絡変更部15と、周波数変換関数作成部16と、スペクトル包絡減算部17と、ダウンサンプリング帯域スペクトル包絡補正部32と、スペクトル包絡減算傾き算出部19と、ダウンサンプリング帯域スペクトル包絡傾き補正部33と、逆フーリエ変換部21と、フレーム合成部22とを具備している。

【0055】

ダウンサンプリング部31は、フレーム切出部11から、入力された音声信号  $x_i(n)$  から切り出されたデータを取得すると、標本化周波数  $f_{s_d} = f_s/4 (= 12 \text{ kHz})$  にダウンサンプリングされたデータ (すなわち、データ長は256サンプル) を算出するように構成されている。

30

【0056】

線形予測分析部13は、ダウンサンプリング部31から取得したダウンサンプリングされたデータに基づく線形予測分析により、ダウンサンプリングされた帯域部分 (音声帯域部分) のみを対象とした線形予測係数  $a_i (i = 1, \dots, p)$  を算出するように構成されている。本実施形態の例では、線形予測係数の次数 (線形予測分析次数)  $p$  を12次とする。

【0057】

例えば、ダウンサンプリング部31を通さず、直接、フレーム切出部11から線形予測分析部13にて線形予測係数を得る方法として、線形予測分析部13は、参考文献2 (コロナ社、「音声の線形予測」、J.D.マーケル、A.H.グレイJr.著、鈴木久喜訳) に記載されている「Selective Linear Prediction (選択線形予測)」方法を用いて、上述の線形予測係数を得るように構成されていてもよい。

40

【0058】

スペクトル包絡算出部14は、線形予測分析部13から線形予測係数を取得すると、スペクトル包絡



【数 2】

$$H(z) = \log \left[ 1 / \left( 1 + \sum_{i=1}^p a_i \cdot z^{-i} \right) \right] \quad (\text{ここでは、複素対数をとっている})$$

を算出するように構成されている。本実施形態の例では、スペクトル包絡のデータ長は、周波数特性の FFT サンプル数の 1/4 である 512 ポイントである。

【0059】

10

ダウンサンプリング帯域スペクトル包絡補正部 32 は、フーリエ変換部 21 から周波数特性  $X(z)$  を取得すると共に、スペクトル包絡減算部 17 からダウンサンプリングされた帯域部分のスペクトル包絡同士を減算することによって得られた補正值  $G(z)$  を取得すると、周波数特性  $X(z)$  のダウンサンプリングされた帯域部分に対して補正值  $G(z)$  の複素指数を取ったものを乗算することにより、補正後の周波数特性

【数 3】

$$X'(z) = X(z) \cdot \exp(G(z)), z = e^{-j2\pi(k-1)/N} \quad (k=1, \dots, N/4, N-N/4+1, \dots, N)$$

20

を取得するように構成されている。

【0060】

ダウンサンプリング帯域スペクトル包絡傾き補正部 33 は、ダウンサンプリング帯域スペクトル包絡補正部 32 から取得した補正後の周波数特性  $X'(z)$  に対して、スペクトル包絡減算傾き算出部 19 から取得した傾き  $G'(z)$  の複素指数をとったものを乗算することにより、補正後の周波数特性の傾き

【数 4】

30

$$X''(z) = X'(z) \cdot \exp(G'(z)), z = e^{-j2\pi(k-1)/N} \quad (k=1, \dots, N/4, N-N/4+1, \dots, N)$$

を算出するように構成されている。

【0061】

逆フーリエ変換部 21 は、ダウンサンプリング帯域スペクトル包絡減算傾き補正部 33 から取得された補正後の周波数特性の傾き  $X''(z)$  に対して逆フーリエ変換 (IFFT) を施すことによって、フレームごとに声質変換された音声信号の時間波形  $x'(m, n)$  を算出するように構成されている。

40

【0062】

ここで、図 6 (a) に、周波数変換倍率 を 0.5 とした場合に本実施形態に係る信号処理装置 1 によって声質が変換された音声信号の音声帯域部分におけるスペクトル包絡を示し、図 6 (b) に、周波数変換倍率 を 0.5 とした場合に本実施形態に係る信号処理装置 1 によって声質が変換された音声信号の音声帯域部分におけるパワースペクトルを示す。

【0063】

50

一方、図7(a)に、周波数変換倍率を2.0とした場合に本実施形態に係る信号処理装置1によって声質が変換された音声信号の音声帯域部分におけるスペクトル包絡を示し、図7(b)に、周波数変換倍率を2.0とした場合に本実施形態に係る信号処理装置1によって声質が変換された音声信号の音声帯域部分におけるパワースペクトルを示す。

#### 【0064】

本実施形態の特徴は、入力された音声信号 $x_i(n)$ に対して声質変換処理(所定処理)を施すように構成されている信号処理装置1であって、音声信号 $x_i(n)$ から周波数特性 $X(z)$ を算出するように構成されているフーリエ変換部(周波数特性算出部)12と、音声信号 $x_i(n)$ から算出された線形予測係数 $a_i$ に基づいて、音声信号 $x_i(n)$ のスペクトル包絡 $H(z)$ を算出するように構成されているスペクトル包絡算出部14と、入力された周波数 $f_{in}$ を一意に変換する周波数変換関数を用いて、スペクトル包絡 $H(z)$ を構成する周波数成分を変更するように構成されているスペクトル包絡変更部(変更部)15と、周波数成分が変更されたスペクトル包絡 $H'(z)$ に基づいて、周波数特性 $X(z)$ を補正するように構成されているダウンサンプリング帯域スペクトル包絡補正部32及びダウンサンプリング帯域スペクトル包絡傾き補正部33(補正部)と、補正された周波数特性 $X''(z)$ に基づいて、声質変換処理が施された音声信号 $x_o(n)$ を取得するように構成されている逆フーリエ変換部21及びフレーム合成部22(処理部)とを具備することを要旨とする。

#### 【0065】

ここで、かかる本実施形態の特徴において、スペクトル包絡算出部14は、上述の線形予測係数 $a_i$ として、入力された音声信号 $x_i(n)$ をダウンサンプリングすることによって得られた音声帯域部分のみを対象とした線形予測係数 $a_i$ を用いるように構成されている。

#### 【0066】

本実施形態に係る信号処理装置1によれば、ダウンサンプリングされた帯域部分において周波数変換関数を利用して声質変換を行うように構成されているため、どのような波形の音声信号が入力された場合であっても、少ない処理量で適切な声質変換を行うことができる。

#### 【0067】

(第3の実施形態)

以下、図8～図10を参照して、本発明の第3の実施形態に係る信号処理装置1について、上述の第2の実施形態に係る信号処理装置1との相違点に着目して説明する。

#### 【0068】

本実施形態に係る信号処理装置1は、入力された音声信号 $x_i(n)$ に対する声質変換処理を行うに際して、上述の第2の実施形態に係る信号処理装置1とは異なり、周波数変換関数の代わりに、ホルマント変更を利用するように構成されている。

#### 【0069】

図8に示すように、本実施形態に係る信号処理装置1は、フレーム切出部11と、フーリエ変換部12と、ダウンサンプリング部31と、線形予測分析部13と、求根部41と、ホルマント算出部42と、ホルマント変更部43と、多項式再構成部44と、スペクトル包絡算出部14と、スペクトル包絡算出部45と、スペクトル包絡減算部17と、ダウンサンプリング帯域スペクトル包絡補正部32と、スペクトル包絡減算傾き算出部19と、ダウンサンプリング帯域スペクトル包絡傾き補正部33と、逆フーリエ変換部21と、フレーム合成部22とを具備している。

#### 【0070】

求根部41は、線形予測分析部13から取得した線形予測係数に基づいて、多項式

【数5】

$$1 + \sum_{i=1}^p a_i \cdot z^{-i} = 0$$

の求根により、複素根としてスペクトル包絡の極  $z_i$  ( $i = 1, \dots, p$ ) を求めるように構成されている。

【0071】

10

ホルマント算出部42は、求根部41から取得したスペクトル包絡の極  $z_i$  に基づいて、共振周波数(位相)  $F_i (= \arg z_i \cdot fs / 2\pi)$  及び帯域幅(振幅)  $B_i (= \log |z_i| \cdot fs / 2\pi)$  を算出するように構成されている。

【0072】

ここで、本実施形態では、共振周波数  $F_i$  がホルマント周波数の存在範囲(およそピッチ周波数以上の200~5000Hz)内にある共振周波数  $F_i$  及び帯域幅  $B_i$  の組み合わせのうち、帯域幅  $B_i$  と共振周波数  $F_i$  との比  $B_i / F_i$  が小さいものを「ホルマント  $F_i, B_i$ 」とみなす。

【0073】

ホルマント変更部43は、ホルマント算出部42から取得したホルマント  $F_i, B_i$  のうち、共振周波数  $F_i$  の低いものから3つまでに対して、ホルマント制御倍率を乗ずることによって、ホルマント  $F_i, B_i$  を変更するように構成されている。

20

【0074】

ここで、共振周波数に対するホルマント制御倍率 及び帯域幅に対するホルマント制御倍率 は、それぞれ独立のものとする。

【0075】

多項式再構成部44は、ホルマント変更部43から取得した変更前のホルマント  $F_i, B_i$  及び変更後のホルマント  $F'_i, B'_i$  に基づいて、複素根  $z'_i$  ( $i = 1, \dots, P$ ) を算出し、

【数6】

30

$$\prod_{i=1}^P (z - z'_i) = z^P + a_i \cdot z^{P-1} + \dots + a_{P-1}' \cdot z + a_P'$$

の関係から、変更後の線形予測係数  $a'_i$  ( $i = 1, \dots, P$ ) を算出するように構成されている。

【0076】

スペクトル包絡算出部45は、多項式再構成部44から取得した変更後の線形予測係数  $a'_i$  に基づいて、変更後のスペクトル包絡

40

【数7】

$$H'(z) = \log \left[ 1 / \left( 1 + \sum_{i=1}^P a'_i \cdot z^{-i} \right) \right]$$

を算出するように構成されている。

【0077】

50

スペクトル包絡減算部 17 は、スペクトル包絡算出部 14 から取得したスペクトル包絡  $H(z)$  及びスペクトル包絡変更部 45 から取得した変更後のスペクトル包絡  $H'(z)$  に基づいて、補正值  $G(z) = H'(z) - H(z)$  を算出するように構成されている。

【0078】

ここで、図9(a)に、共振周波数に対するホルマント制御倍率を0.5とした場合に本実施形態に係る信号処理装置1によって声質が変換された音声信号の音声帯域部分におけるスペクトル包絡を示し、図9(b)に、共振周波数に対するホルマント制御倍率を0.5とした場合に本実施形態に係る信号処理装置1によって声質が変換された音声信号の音声帯域部分におけるパワースペクトルを示す。

【0079】

一方、図10(a)に、共振周波数に対するホルマント制御倍率を2.0とした場合に本実施形態に係る信号処理装置1によって声質が変換された音声信号の音声帯域部分におけるスペクトル包絡を示し、図10(b)に、共振周波数に対するホルマント制御倍率を2.0とした場合に本実施形態に係る信号処理装置1によって声質が変換された音声信号の音声帯域部分におけるパワースペクトルを示す。

【0080】

本実施形態の特徴は、入力された音声信号  $x_i(n)$  に対して声質変換処理(所定処理)を施すように構成されている信号処理装置1であって、音声信号  $x_i(n)$  から周波数特性  $X(z)$  を算出するように構成されているフーリエ変換部(周波数特性算出部)12と、音声信号  $x_i(n)$  から算出された線形予測係数  $a_i$  に基づいて、音声信号  $x_i(n)$  のスペクトル包絡  $H(z)$  を算出するように構成されているスペクトル包絡算出部14と、ホルマント  $F_i, B_i$  に所望の変更を加えることによって(ホルマント制御倍率を乗ずることによって)、スペクトル包絡  $H(z)$  に対して所望の変更を加えるように構成されているホルマント変更部43及びスペクトル包絡変更部45(変更部)と、所望の変更が加えられたスペクトル包絡  $H'(z)$  に基づいて、周波数特性  $X(z)$  を補正するように構成されているダウンサンプリング帯域スペクトル包絡補正部32及びダウンサンプリング帯域スペクトル包絡傾き補正部33(補正部)と、補正された周波数特性  $X''(z)$  に基づいて、声質変換処理が施された音声信号  $x_o(n)$  を取得するように構成されている逆フーリエ変換部21及びフレーム合成部22(処理部)とを具備することを要旨とする。

【0081】

ここで、かかる本実施形態の特徴において、スペクトル包絡算出部14は、上述の線形予測係数  $a_i$  として、音声帯域部分(ダウンサンプリングされた帯域部分)のみを対象とした線形予測係数  $a_i$  を用いるように構成されている。

【0082】

本実施形態に係る信号処理装置1によれば、ダウンサンプリングされた帯域部分においてホルマント変更を利用して声質変換を行うように構成されているため、ホルマント周波数に対応した細かい制御による高品質な声質変換を行うことができる。

【0083】

(その他の実施形態)

上述のように、本発明について、上述した第1~第3の実施形態によって説明したが、かかる実施形態における開示の一部をなす論述及び図面は、本発明を限定するものであると理解すべきではない。かかる開示から当業者には様々な代替実施形態、実施例及び運用技術が明らかとなる。

【0084】

また、上述の実施形態では特に触れていないが、上述の信号処理装置1によって行われる各処理をコンピュータに実行させるプログラムが提供されてもよい。また、かかるプログラムは、コンピュータ読取り可能媒体に記録されていてもよい。コンピュータ読取り可能媒体を用いれば、かかるプログラムをコンピュータにインストールすることが可能である。ここで、かかるプログラムが記録されたコンピュータ読取り可能媒体は、非一過性の

10

20

30

40

50

記録媒体であってもよい。非一過性の記録媒体は、特に限定されるものではないが、例えば、CD-ROMやDVD-ROM等の記録媒体であってもよい。

【0085】

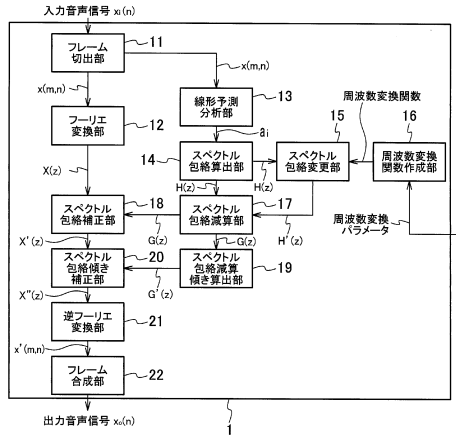
或いは、上述の信号処理装置1によって行われる各処理を実行するためのプログラムを記憶するメモリ及びメモリに記憶されたプログラムを実行するプロセッサによって構成されるチップが提供されてもよい。

【符号の説明】

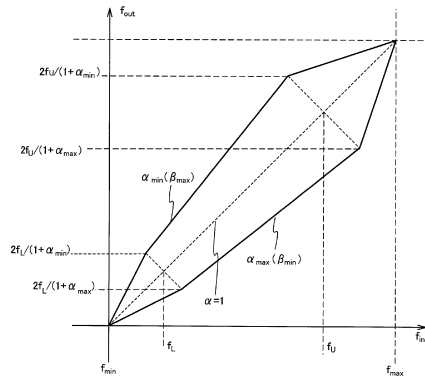
【0086】

- 1 ... 信号処理装置
- 1 1 ... フレーム切出部 10
- 1 2 ... フーリエ変換部
- 1 3 ... 線形予測分析部
- 1 4、4 5 ... スペクトル包絡算出部
- 1 5 ... スペクトル包絡変更部
- 1 6 ... 周波数変換関数作成部
- 1 7 ... スペクトル包絡減算部
- 1 8 ... スペクトル包絡補正部
- 1 9 ... スペクトル包絡減算傾き算出部
- 2 0 ... スペクトル包絡傾き補正部
- 2 1 ... 逆フーリエ変換部 20
- 2 2 ... フレーム合成部
- 3 1 ... ダウンサンプリング部
- 3 2 ... ダウンサンプリング帯域スペクトル包絡補正部
- 3 3 ... ダウンサンプリング帯域スペクトル包絡傾き補正部
- 4 1 ... 求根部
- 4 2 ... ホルマント算出部
- 4 3 ... ホルマント変更部
- 4 4 ... 多項式再構成部

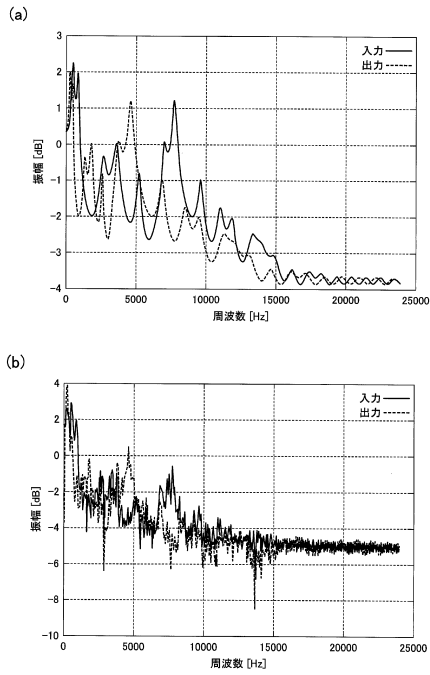
【図1】



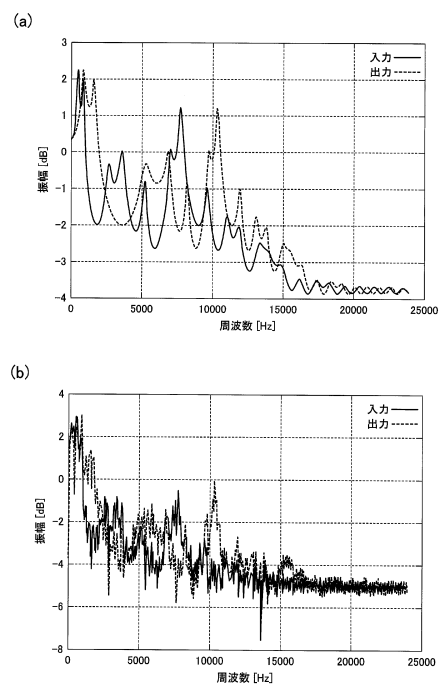
【図2】



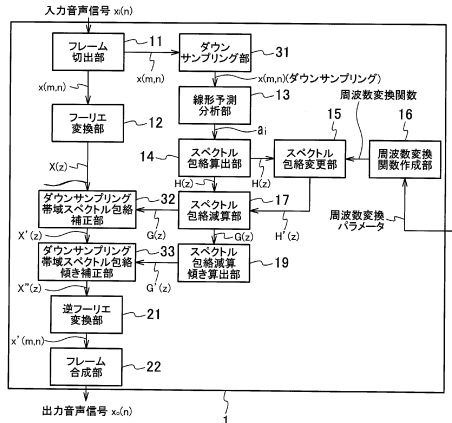
【図3】



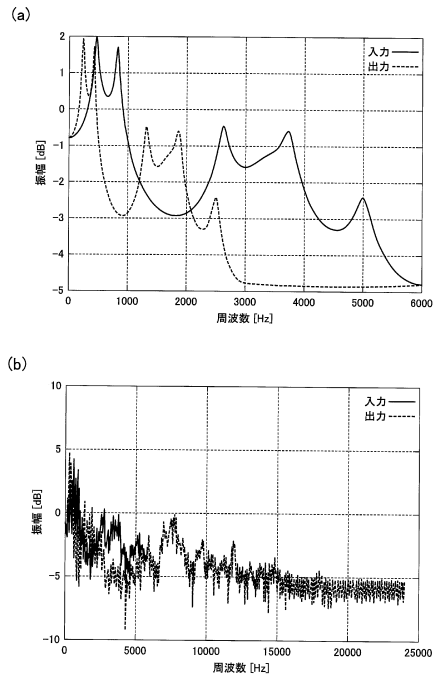
【図4】



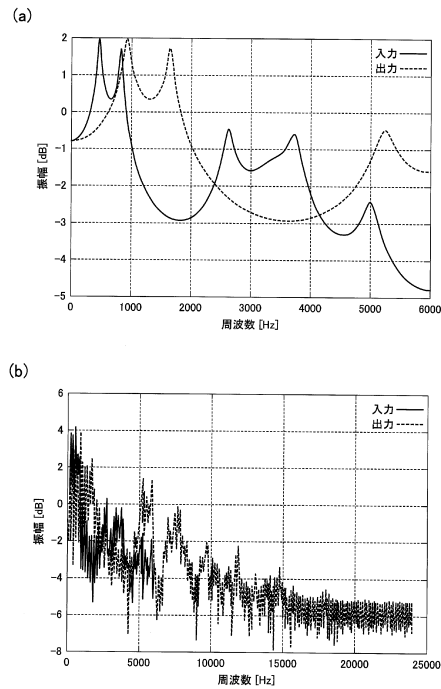
【図5】



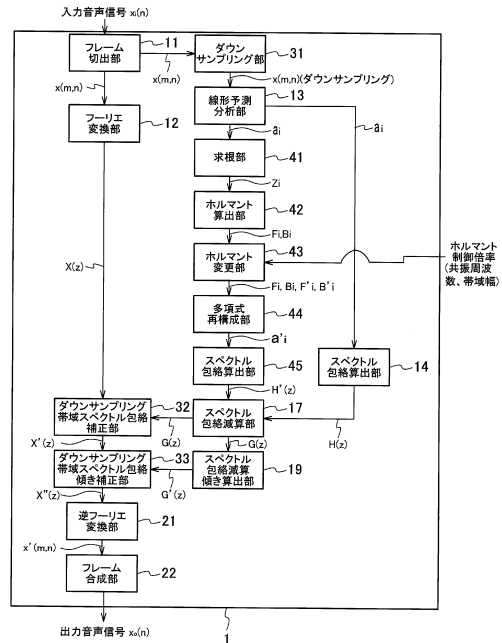
【図6】



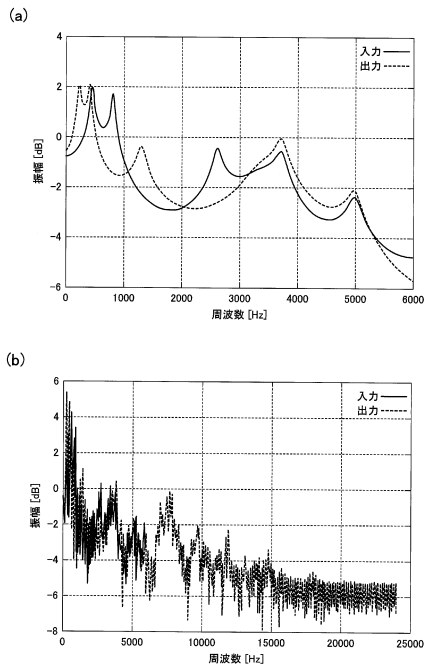
【図7】



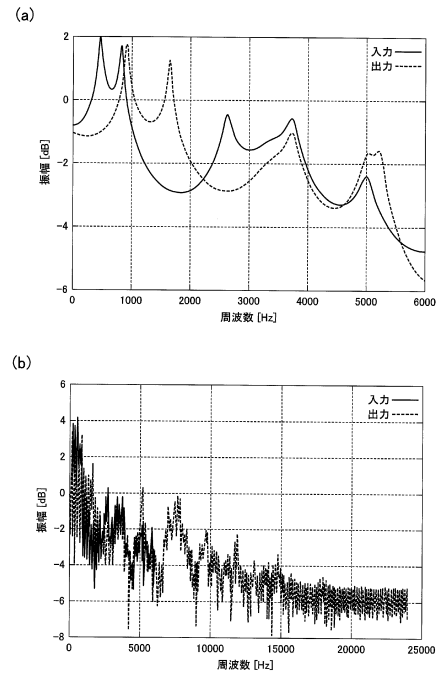
【図8】



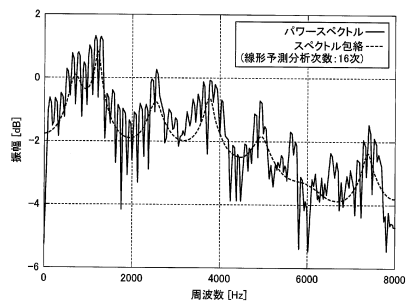
【図 9】



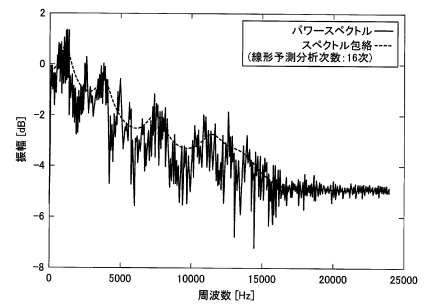
【図 10】



【図 11】

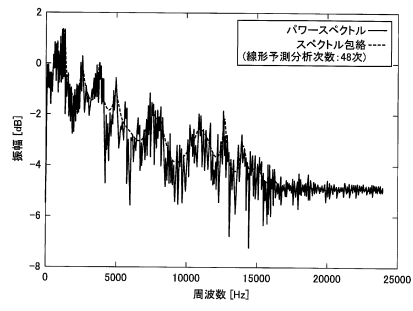


【図 12】





【図13】



---

フロントページの続き

(72)発明者 都木 徹

東京都世田谷区砧一丁目10番11号 一般財団法人NHKエンジニアリングシステム内

審査官 大野 弘

(56)参考文献 特開平06-110499(JP,A)

特表2001-522471(JP,A)

(58)調査した分野(Int.Cl., DB名)

G10L 21/007

G10L 25/12

G10L 25/15

G10L 25/18