

(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2021-157128
(P2021-157128A)

(43) 公開日 令和3年10月7日(2021.10.7)

(51) Int.Cl.

F I

テーマコード (参考)

G 1 O L 19/018 (2013.01)

G 1 O L 19/02 (2013.01)

G 1 O L 19/018

G 1 O L 19/02 1 7 O Z

審査請求 未請求 請求項の数 9 O L (全 13 頁)

(21) 出願番号	特願2020-59682 (P2020-59682)	(71) 出願人	000208891
(22) 出願日	令和2年3月30日 (2020.3.30)		K D D I 株式会社
			東京都新宿区西新宿二丁目3番2号
		(74) 代理人	100092772
			弁理士 阪本 清孝
		(74) 代理人	100119688
			弁理士 田邊 壽二
		(72) 発明者	西澤 信行
			埼玉県ふじみ野市大原二丁目1番15号
			株式会社K D D I 総合研究所内

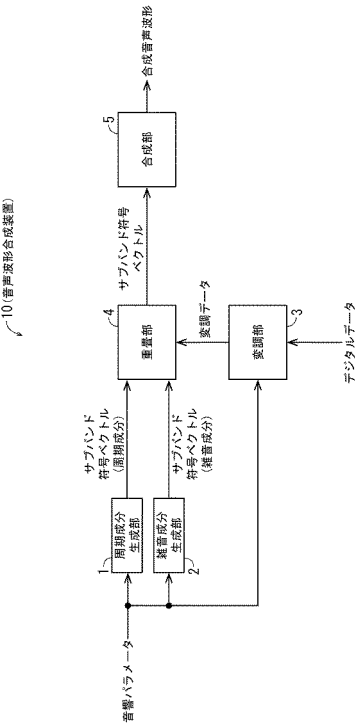
(54) 【発明の名称】 音声波形合成装置、方法及びプログラム

(57) 【要約】

【課題】 デジタルデータが重畳された音声を高速に合成することのできる音声波形合成装置を提供する。

【解決手段】 スペクトル包絡特性を含む音響パラメータを入力として、フィルタバンクにより分割された複数の周波数帯域であるサブバンドにおいて音声波形を合成する音声波形合成装置10であって、前記スペクトル包絡特性を参照して第1のサブバンド符号ベクトルを生成し、重畳対象となるデジタルデータをデジタル変調し、前記デジタル変調した結果を、前記スペクトル包絡特性を参照して当該デジタル変調した結果の振幅を調整してから前記第1のサブバンド符号化ベクトルの所定の要素に対して重畳することによって第2のサブバンド符号化ベクトルを生成し、前記第2のサブバンド符号化ベクトルを復号することによって合成音声波形を合成することを特徴とする。

【選択図】 図1



【特許請求の範囲】**【請求項 1】**

スペクトル包絡特性を含む音響パラメータを入力として、フィルタバンクにより分割された複数の周波数帯域であるサブバンドにおいて音声波形を合成する音声波形合成装置であって、

前記スペクトル包絡特性を参照して第 1 のサブバンド符号ベクトルを生成し、

音声に重畳するデジタルデータをデジタル変調し、前記デジタル変調した結果を、前記スペクトル包絡特性を参照して当該デジタル変調した結果の振幅を調整してから前記第 1 のサブバンド符号ベクトルの所定の要素に対して重畳することによって第 2 のサブバンド符号ベクトルを生成し、

前記第 2 のサブバンド符号ベクトルを復号することによって合成音声波形を合成することを特徴とする音声波形合成装置。

【請求項 2】

前記デジタル変調することは、前記フィルタバンクを構成するバンドパスフィルタの中心周波数を、当該デジタル変調の搬送波の周波数とすることによって実行されることを特徴とする請求項 1 に記載の音声波形合成装置。

【請求項 3】

前記重畳することは、前記第 1 のサブバンド符号ベクトルの雑音成分の全部又は一部を、当該雑音成分のパワーと同じパワーを有するデジタル変調成分で置き換えることによって実行されることを特徴とする請求項 1 または請求項 2 に記載の音声波形合成装置。

【請求項 4】

前記デジタル変調することは、前記第 1 のサブバンド符号ベクトルの正弦波成分及び雑音成分のパワーと比較して所定割合で小さいパワーを有するデジタル変調成分を生成することによって実行され、

前記重畳することは、当該所定割合で小さいパワーを有するデジタル変調成分を重畳することによって実行されることを特徴とする請求項 1 または請求項 2 に記載の音声波形合成装置。

【請求項 5】

前記デジタル変調された成分を重畳するサブバンドが、時間軸上において変更されるように設定されていることを特徴とする請求項 1 ないし 4 のいずれかに記載の音声波形合成装置。

【請求項 6】

前記重畳対象となるデジタルデータには、いずれのサブバンドにデジタルデータが重畳されているかを指定する情報が含まれ、当該指定する情報の前記デジタル変調された成分が、所定のサブバンドに対して重畳されることを特徴とする請求項 1 ないし 5 のいずれかに記載の音声波形合成装置。

【請求項 7】

前記重畳対象となるデジタルデータには、スペクトル包絡特性を指定する情報が含まれ、

当該スペクトル包絡特性を指定する情報の前記デジタル変調された成分が、所定のサブバンドに対して重畳され、

スペクトル包絡特性に応じて、いずれのサブバンドにデジタルデータを重畳するかの所定の条件が予め設定されていることを特徴とする請求項 1 ないし 6 のいずれかに記載の音声波形合成装置。

【請求項 8】

スペクトル包絡特性を含む音響パラメータを入力として、フィルタバンクにより分割された複数の周波数帯域であるサブバンドにおいて音声波形を合成する音声波形合成方法であって、

前記スペクトル包絡特性を参照して第 1 のサブバンド符号ベクトルを生成し、

音声に重畳するデジタルデータをデジタル変調し、前記デジタル変調した結果を、前記

10

20

30

40

50

スペクトル包絡特性を参照して当該デジタル変調した結果の振幅を調整してから前記第 1 のサブバンド符号ベクトルの所定の要素に対して重畳することによって第 2 のサブバンド符号ベクトルを生成し、

前記第 2 のサブバンド符号ベクトルを復号することによって合成波形音声波形を合成することを特徴とする音声波形合成方法。

【請求項 9】

コンピュータを請求項 1 ないし 7 のいずれかに記載の音声波形合成装置として機能させることを特徴とするプログラム。

【発明の詳細な説明】

【技術分野】

10

【0001】

本発明は、デジタルデータを重畳して音声波形の合成を行う音声波形合成装置、方法及びプログラムに関する。

【背景技術】

【0002】

音声合成技術とは音声を人工的に合成する手法である。代表的な利用方法として、テキスト音声変換 (Text-To-Speech) が挙げられるが、例えば日本語では、TTS の入力となるテキストは通常、漢字仮名交じり文であり、例えば文字と合成すべき音声の特徴とを直接マッピングすることはその関係性の構造が極めて複雑であることから困難である。そこで音素や韻律情報 (音素列内でのアクセント核位置やアクセント句境界位置) 等といった抽象化された形態により音声に関する特徴を記述した中間表現を用い、まずテキストから中間表現、中間表現から音声の音響的 (物理的) 特徴、という 2 段階の変換を経て、音声の音響的特徴の情報に合う音声波形を信号処理的に生成する、あるいは事前準備した波形の蓄積から適切なものを選択することで、合成音声波形を得ることができる。以下ではこの最後の段階の処理、例えばテキスト音声変換システムにおいて、音声の音響的特徴から、信号処理的に音声波形を合成する機構を対象とする。

20

【0003】

以下では、そのような機構を音声波形合成装置という。すなわち、以下において音声波形合成装置とは、テキスト音声変換システムのうち、先述の 2 段階の変換を経た後の処理である、音声波形の合成処理を信号処理的に行う機構に相当するものをいう。

30

【0004】

音声波形合成装置の入力は、音声の音響的特徴を表す音響パラメータの時系列データであり、例えば 5 ms 間隔のケプストラム係数や基本周波数の値である。より抽象的にはケプストラム係数や基本周波数の値をその要素とするベクトルの、時系列データである。また出力は音声波形に対応するデジタルデータで、例えば PCM (パルス符号変調) データである。このような音声波形合成装置は、テキスト音声変換システムの後段側の構成として利用される以外にも、音声符号化システムの復号部にも含まれ、利用されるものである。

【0005】

このような音声波形合成装置を実現する方法の 1 つに、特許文献 1 に開示されるサブバンド符号化技術を応用した音声波形生成処理がある。この方法では、仮想的に複数の正弦波を足し合わせることで音声波形のうちの周期性成分を合成するが、まず各正弦波の成分に対応するサブバンド符号 (ベクトル) を計算により求め、その振幅を音声のスペクトル包絡特性に合わせて変更してから各正弦波に対応するサブバンド符号をサブバンド符号上で足し合わせる。また、非周期性成分についても疑似雑音系列の振幅について、各サブバンドの非周期性成分の例えば平均二乗振幅が一致するように修正し、それらを要素とするサブバンド符号を求める。最終的に、周期性成分のサブバンド符号と、非周期性成分サブバンド符号を足した結果を合成音声のサブバンド符号とし、それを復号することで音声波形を合成する。

40

【0006】

50

サブバンド符号の帯域分割数がMであるとき、サブバンド符号化ではフィルタバンクによる帯域分割後、通常、M：1のサンプル間引きを行う（最大間引き）。サブバンド符号のベクトルの次元数もMなので、符号化前後でその値の数の合計は変わらないものの、正弦波をサブバンド符号化すると、符号化結果のベクトルは近似的にスパースになる（ベクトルの要素の多くが0と見なせる小さい値になる）ので、正弦波の振幅を乗算処理によって変更する際、0と見なせる要素に対する乗算処理を省くことができ、処理量を大幅に削減できる。また、定常な正弦波をサブバンド符号化した結果のサブバンド符号ベクトルは、波形一般を対象とする、フィルタバンクを用いた（通常の）サブバンド符号化器を用いなくても簡単な計算で求められ、非周期性成分も疑似雑音系列の振幅を変更するだけで生成できる。従って、最終的に必要なサブバンド符号の復号処理を考慮しても、全体として高速な音声波形生成が可能である。

10

【0007】

一方、音響信号に対して、音響信号とは別のデジタルデータを重畳する方法が知られており、非特許文献1に開示されている。例えば合成音声に対してこの方法で追加の情報を重畳することで音声合成技術の利便性を高めることができる。

【先行技術文献】

【特許文献】

【0008】

【特許文献1】特開2014-109667号公報

【非特許文献】

20

【0009】

【非特許文献1】松岡 保静, "音響データ通信技術：音響OFDM(<小特集>携帯情報機器における音響技術)", "日本音響学会誌 68(3), 143-147, 2012

【発明の概要】

【発明が解決しようとする課題】

【0010】

従来、合成音声の波形に対して音響的にデジタルデータを重畳する場合、重畳した結果生じる聴感的な影響を抑えるためには、合成音声波形を音響分析し、分析した結果にデジタルデータを埋め込んだ上で再度音声波形合成の処理を行う必要があり、その処理量が大幅に増加するという課題があった。特に先述のサブバンド符号化技術を応用した音声波形生成方式は低速なプロセッサでも実用的な処理速度が得られるという特長を有するが、合成した音声に対してデジタルデータをさらに重畳する場合、重畳に必要な処理量が大きいために、システムとしては、より高速なプロセッサ等を用いる必要が生じるといった課題があった。

30

【0011】

上記従来技術の課題に鑑み、本発明は、デジタルデータが重畳された音声を高速に合成することのできる音声波形合成装置、方法及びプログラムを提供することを目的とする。

【課題を解決するための手段】

【0012】

上記目的を達成するため、本発明は、スペクトル包絡特性を含む音響パラメータを入力として、フィルタバンクにより分割された複数の周波数帯域であるサブバンドにおいて音声波形を合成する音声波形合成装置であって、前記スペクトル包絡特性を参照して第1のサブバンド符号ベクトルを生成し、音声に重畳するデジタルデータをデジタル変調し、前記デジタル変調した結果を、前記スペクトル包絡特性を参照して当該デジタル変調した結果の振幅を調整してから前記第1のサブバンド符号ベクトルの所定の要素に対して重畳することによって第2のサブバンド符号ベクトルを生成し、前記第2のサブバンド符号ベクトルを復号することによって合成音声波形を合成することを特徴とする。また、当該音声波形合成装置に対応する方法又はプログラムであることを特徴とする。

40

【発明の効果】

【0013】

50

本発明によれば、サブバンド符号上においてデジタルデータを重畳することにより、デジタルデータが重畳された音声を高速に合成することができる。

【図面の簡単な説明】

【0014】

【図1】一実施形態に係る音声波形合成装置の機能ブロック図である。

【図2】本実施形態における周期成分生成部、雑音成分生成部、変調部及び重畳部における処理を説明するための図である。

【図3】本実施形態におけるデジタルデータが重畳された合成音声を、模式例なスペクトルとして示す図である。

【図4】一般的なコンピュータにおけるハードウェア構成を示す図である。

10

【発明を実施するための形態】

【0015】

図1は、一実施形態に係る音声波形合成装置10の機能ブロック図である。音声波形合成装置10は、周期成分生成部1、雑音成分生成部2、変調部3、重畳部4及び合成部5を備える。音声波形合成装置10は、その全体的な動作として、例えば1ミリ秒～20ミリ秒程度といったような音声波形が定常とみなせる所定間隔ごとの音響パラメータ（基本周波数、非周期性情報及びスペクトル包絡特性を含む）と、当該所定間隔における重畳対象となるデジタルデータと、を入力として受け取り、当該デジタルデータが重畳された合成音声波形を出力するものである。なお、音声波形合成装置10から変調部3を省略し、デジタルデータを入力として用いず、合成音声にデジタルデータを重畳しない場合の構成は、特許文献1の構成（正弦波の足し合わせによる周期性成分と、雑音成分とで音声波形の合成を行う場合の構成）と同様である。

20

【0016】

周期成分生成部1は、音響パラメータとしての基本周波数、非周期性情報およびスペクトル包絡特性を参照することによって、音声の周期性成分である各調波成分（基本周波数の整数倍の周波数の正弦波成分）の振幅を求め、各調波の成分の和によって表現される、音声の周期性成分に関するM次元のサブバンド符号ベクトルを生成し（サンプリングレートは音声波形の1/Mである）、重畳部4へと出力する。雑音成分生成部2は、内部的に複数の周波数バンドで合成され、各バンドは帯域制限された白色雑音（ここでは、白色雑音にバンドパスフィルタを適用したものをそう呼ぶ。以下同様。）源を有し、非周期性情報とスペクトル包絡特性を参照することによって各バンドの振幅調整を行って、音声の非周期性成分に対応する雑音成分を生成し、重畳部4へ出力する。ただし実際の雑音成分は、周期成分生成部1と同様に、M次元のサブバンド符号ベクトルの形で生成される。なお、周期成分生成部1では、当該周波数における非周期性の値が大きいほど、その周波数の正弦波の振幅を小さくし、雑音成分生成部2では、当該周波数における非周期性の値が大きいほど（仮に全て雑音である場合の値を1と定義するのではあれば、値が1に近くなるほど）、スペクトル包絡特性に近いパワーとなるように振幅を調整する。これは、デジタルデータの重畳を行わない、従来の音声波形合成装置でも同様である。

30

【0017】

周期成分生成部1及び雑音成分生成部2のサブバンド符号化処理は、特許文献1と同様でよく、入力信号をバンドパスフィルタによって複数（M個）のサブバンドに分割し、1/Mのサンプリング周波数になるようにダウンサンプリングしてから、所定の時刻における各サブバンドの値を並べることでベクトルを作り、さらに、スペクトル包絡特性を参照して振幅調整のための値を乗じたものを、サブバンド符号ベクトルとして出力すればよい。ただし、上記処理は説明のための仮想的な処理であり、特許文献1に示されるように、フィルタバンクを用いたサブバンド符号化器を使わなくても、計算でサブバンド符号ベクトルをそれぞれ求めることができる。

40

【0018】

変調部3は、重畳対象となるデジタルデータを変調して、重畳部4へと出力する。変調部3の詳細は後述するが、内部は1以上のバンドで構成され、各バンドがそれぞれデジタル

50

変調器を持ち、周期成分生成部1や雑音成分生成部2の出力と同様に、最終的に合成する音声の $1/M$ のサンプリングレートでの変調結果を出力する。また、周期成分生成部1及び雑音成分生成部2の処理でも用いるスペクトル包絡特性や、基本周波数、非周期性情報といった情報（音響パラメータの情報）を変調部3でも入力として受け取ることで共有しておき、それらを参照して各バンドでの変調結果に振幅調整係数を乗じることでバンド毎にそのパワーの調整された変調データを得る。後述するように、変調部3におけるデジタル変調には、既存のデジタル変調方式を用いればよい。

【0019】

重畳部4では、周期成分生成部1及び雑音成分生成部2で得られたサブバンド符号ベクトルの和（周期成分及び雑音成分を有し、以下、これを第1のサブバンド符号ベクトルと呼ぶ）に対して、変調部3から得られた、最終的に合成する音声の $1/M$ のサンプリングレートで出力される、複数バンドの変調データを、第1のサブバンド符号ベクトルの所定のサブバンドの要素に対して重畳することにより、デジタルデータが重畳された第2のサブバンド符号ベクトルを生成し、合成部5へと出力する。なお、第2のサブバンド符号ベクトルと第1のサブバンド符号ベクトルとの差は、計算誤差等を考えなければ、実際に重畳される変調後のデジタルデータの波形を、サブバンド符号化した結果と等しい。

【0020】

合成部5は、重畳部4で得た第2のサブバンド符号ベクトルを復号することにより、デジタルデータが重畳された合成音声波形を出力する。合成部5による合成処理も、特許文献1と同様でよい。

【0021】

図2は、本実施形態における周期成分生成部1、雑音成分生成部2、変調部3及び重畳部4における処理を説明するための図である。図2において、上段部分は周期成分生成部1が生成する K 個の正弦波成分を表しており、ベクトル加算処理V1により周期成分のサブバンド符号ベクトルが出力される。中段部分は、雑音成分生成部2で生成する N バンドの帯域制限された白色雑音系列 $W_1, \dots, W_{N-1}$ を表しており、 $b_1, \dots, b_{N-1}$ は N バンドの系列それぞれに対する振幅調整係数であり、ベクトル加算処理V2により雑音成分のサブバンド符号ベクトルが出力される。（なお、図2にも示されるようにベクトル加算処理V1及びV2はそれぞれ、周期成分生成部1及び雑音成分生成部2における振幅調整後の処理となる。）前述したように、ベクトル加算処理V1の出力とベクトル加算処理V2の出力とは、重畳部4での処理に含まれる加算処理Aにより加算されることで、第1のサブバンド符号ベクトルが得られる。下段部分の $b'_1, \dots, b'_{N-1}$ は変調部3（デジタル変調器 $M_1, \dots, M_{N-1}$ ）による変調結果に対する振幅調整係数である。説明を簡単にするために、変調部3は、雑音成分生成部2におけるバンド構成と同様のバンド構成（同じフィルタバンクを用いる構成）とし、そのそれぞれにおいて、デジタル変調を行う構成を考える。ただし後述するように、例えばデジタルデータの重畳前後で、音声のスペクトル包絡特性等が聴感上影響のあるレベルで大きく変わらなければよく、雑音成分生成部2と変調部3のバンド構成は必ずしも同一である必要はない。

【0022】

本実施形態では、以下の（1）や（2）の手法により、重畳部4でデジタルデータが重畳されたサブバンド符号ベクトルが生成され、当該重畳された音声のサブバンド符号ベクトルから合成部5において音声波形を合成することができる。なお、図2において、変調部3では複数のバンドでそれぞれデジタル変調器が配置されているが、実際に同時にデジタルデータを重畳するバンドはそのうちの一部だけである。一般に、同時にデジタルデータを重畳するバンド数が増えれば、大量のデータは重畳できるが、聴感上の影響が大きくなる。変調部3でデジタルデータを重畳しないバンドでは、変調器は値0を出力するものとする。なお、これは当該バンドで b'_i を0とするのと等価である。

【0023】

（1）デジタルデータが重畳された音声のスペクトル包絡特性が、元の音声のスペクトル包絡特性を再現するように、雑音成分を同じパワーのデジタル変調結果で置き換える。先

述したように、ここでは雑音成分生成部2と変調部3のバンド構成は同じなので、変調部3の所定のバンドにおいて、同じバンドの雑音成分のパワーを再現するように b'_i を調整し、代わりに雑音成分生成部2における当該バンドの雑音成分を0にする。すなわち、 b_i を0とする。あるいは、聴感上の影響を抑えるために、デジタルデータの復調に影響しない範囲で、 b_i と b'_i を調整し（ b_i を0に近づくほど復号が容易になって大量のデジタルデータの埋込が可能となり、一方で聴感上の影響を抑えるためには、 b_i の値が元の値よりもあまり小さくならないことが望ましい）、雑音成分が少し残る形で変調結果を重畳させても良い。この際、調整した b_i と b'_i とによる、雑音成分生成部2と変調部3の当該バンド出力の和のパワーが、調整前の b_i による雑音成分生成部2の当該バンド出力のパワーと等しくなるよう、調整してよい。

10

【0024】

(2) 周期成分生成部1における正弦波の振幅 a_i 、 a_{i-1} および雑音成分生成部2における雑音系列のパワー b_i 、 b_{i-1} と比較して、その復調に問題のないレベルでデジタル変調結果の振幅調整係数 b'_i 、 b'_{i-1} を小さく設定し（例えばルールベースで所定割合だけ小さい値を設定すればよい）、変調部3において所定のバンドでデジタル変調結果を生成し（それ以外のバンドは値0とする）、この結果を重畳部4において重畳する。これは、デジタル変調結果のパワーが元の音声波形よりも小さければ聴感上への影響も小さいであろうとの考察に基づく。シンボル長(時間長)が十分に長ければ、当該サブバンドにデジタル変調結果以外の成分が含まれていてもデジタルデータの復調は可能となる。

【0025】

なお、(1)、(2)のいずれの場合においても、合成音声からデジタルデータを復号する際は、非特許文献1等の既存手法と同様に、搬送波との相関を取るようにすればよい。

20

【0026】

また、特に低域においては周期成分を、一般的に雑音的な特徴を持つデジタルデータの変調結果で置き換えると聴感上の影響が大きいと考えられることから、(1)では雑音成分とデジタルデータの置き換えに限定したが、例えば高域のように聴感上の影響がより小さいと考えられる場合等、周期成分も考慮する対象に含めることができる。この場合、 b_i 、 b_{i-1} と b'_i 、 b'_{i-1} の間だけでなく、 a_i 、 a_{i-1} も含めてスペクトル包絡特性を再現するように制御すれば良い。例えば、雑音成分生成部2の所定のバンドのパワーと、周期成分合成部1の出力である正弦波の和に対してバンドパスフィルタにより前記所定のバンドと同様の帯域制限を行った結果に対するパワーと、変調部3の同じバンドのパワーと、の3者の和が、デジタルデータ重畳前の、前2者のパワーの和と等しくなるように、 a_i 、 a_{i-1} 、 b_i 、 b_{i-1} 、 b'_i 、 b'_{i-1} を決めればよい。なお、雑音成分生成部2および変調部3のバンド構成が、サブバンド構成と同じである場合については、帯域制限された周期性成分に対するパワーを求めるために、まず、各調波（正弦波）に対するサブバンド符号ベクトルの和を計算して周期性成分全体に対するサブバンド符号ベクトルを求め、その符号ベクトルのうちの、対象のバンドに対応する要素の時系列データから、当該バンドのパワーを求めるといった方法を用いてもよい。

30

【0027】

図3は、本実施形態におけるデジタルデータが重畳された合成音声を、模式例なスペクトルとして示す図である。例EX1は対比例として、デジタルデータを全く重畳しない場合（特許文献1の手法を用いる場合）の合成音声のスペクトルが、破線で示される包絡特性SPと、縦方向の実線で示される調波成分（正弦波成分） $F_0, 2F_0, 3F_0$ とを有するものとして示されている。例EX2は、例EX1の合成音声に関して本実施形態のデジタルデータ重畳を実行した場合のスペクトル例が示されている。元の包絡特性SPのうち、部分Bの箇所はデジタルデータ重畳に用いるため、デジタルデータが重畳される周波数帯域において正弦波の振幅を小さくすることで調波成分のパワーを減らし、同パワーを有しスペクトルdSPを有するデジタル変調成分が重畳されている。

40

【0028】

50

以上、本実施形態の音声波形合成装置10によれば、サブバンド符号化技術に基づく音声波形合成装置において、サブバンド符号上でデジタルデータを埋め込むことにより、デジタルデータをデジタル変調した波形を重畳した、あるいは一部のサブバンドの帯域の波形をデジタル変調されたデータで置き換えた合成音声波形を高速に生成することが可能となる。特に、ダウンサンプリングされたサブバンド符号に対してデジタルデータを重畳することで、デジタルデータを含むような合成音声波形を、音声波形合成とデータ埋め込みを順に処理する場合よりも高速に生成することができる。

【0029】

以下、各種の追加説明等を行う。

【0030】

(A) 音声波形合成装置10でのサブバンド符号化におけるフィルタバンクの設計(及び復号と、デジタル変調及び復調)は、以下の(A-1)~(A-7)のような考察に則って行うようにすればよい。

【0031】

(A-1) 一般に、サブバンド符号化におけるフィルタバンクの設計では、符号化時のダウンサンプリングと、復号時のアップサンプリングに伴い生じるエイリアス成分をそれぞれ打ち消しあうようなフィルタバンクが用いられる。なお、実際のシステムではデータ圧縮のために値の量子化が行われ、その量子化誤差に起因するエイリアスがある。このエイリアスに伴う雑音を考慮しつつ、各サブバンドの量子化誤差の大小を制御することで、聴感上の影響を抑えつつデータの圧縮を行うような処理が行う必要がある。したがって、最大間引きのサブバンド符号に対して情報を埋め込む場合、一般には、それにより生じるエイリアスの影響を考えなければならない。

【0032】

(A-2) ただし、サブバンド符号化では、例えばMPEGオーディオ等で用いられる疑似QMFバンクのように、各サブバンドの中心周波数の成分は全て通過し、かつ、その周波数は他のサブバンドでは完全に遮断されるように設計したフィルタバンクが用いられることが多い。このようなフィルタバンクは、まず、プロトタイプフィルタと呼ばれるローパスフィルタ特性を持つFIRフィルタを設計し、そのFIRフィルタの係数をコサイン関数で変調することでフィルタバンクを設計できる。 $\cos x = 1/2\{(\exp(jx) + \exp(-jx))\}$ (ここでjは虚数単位を表す)であることから、コサイン関数でFIRフィルタの係数を変調する(フィルタ係数のそれぞれにコサイン関数を掛ける)と、その周波数・振幅特性は、プロトタイプフィルタの周波数・振幅特性を周波数軸方向にプラスおよびマイナスにシフトさせたものの和になる。周波数・振幅特性において負の周波数まで考えると、ローパスフィルタは周波数0を中心周波数とするバンドパスフィルタと見なせるので、フィルタ係数に対してコサイン変調を行うことで、その中心周波数がコサイン関数の周波数で決まるバンドパスフィルタが得られる。そして、この周波数を変えて複数のフィルタを構築することで、フィルタバンクを設計できる。

【0033】

M帯域の等分割、最大間引きのフィルタバンクを構成する場合、プロトタイプフィルタは角周波数0で全通過となるように、また、 $\pi/2M$ 以上が遮断されるように設計する。これを満たす、いわゆる理想フィルタ(通過域で振幅特性が1、遮断域で0)を実現するためには、必要なフィルタ長は無限大となり、このようなフィルタは近似的に実現する場合でも必要なフィルタ長が極端に長くなる。そこで、一般には最終的なフィルタバンクでは、あるバンドの通過域が、隣のバンドの中心周波数までオーバーラップするような、フィルタの周波数・振幅特性がなだらかなフィルタを設計する(プロトタイプフィルタにおいては、遮断域のエッジ角周波数を π/M にする)。この場合、各バンドの中心周波数のみが、他のバンドの影響を受けない形になる。なお、フィルタの設計では、最終的に構築するフィルタバンクの全てのバンドの通過特性を足し合わせた結果が、フラット(全通過)となるような基準と合わせてプロトタイプフィルタを設計する。疑似QMFバンクの場合は、このようなフィルタは近似的に実現される。

10

20

30

40

50

【0034】

(A - 3) 従ってサブバンド符号化では、サブバンドの中心周波数の成分のみは、符号化時のダウンサンプリングと復号時のアップサンプリングに伴い生じるエイリアスの影響を受けない形に、あるいは近似的にほぼ無視できる形に設計可能で、実際にMPEGオーディオにおけるサブバンド符号化ではそのような設計のフィルタバンクが用いられている。従って、当該バンドの中心周波数を搬送波の周波数とするようなデジタル変調を各バンドにおいて行えば、復号時のアップサンプリングで生じるエイリアスの影響を受けにくくなる。ここで、「受けにくくなる」とは、実際にはデジタル変調結果で搬送波の振幅や位相特性が時間的に変化することで、実際の周波数スペクトル特性は搬送波の周波数を中心に周波数軸上で広がった形になり、この広がった部分ではエイリアスの影響があるためである。デジタル変調におけるシンボルレートを上げるほどこの広がりは大きくなり、エイリアスの影響がより大きくなる。逆にシンボルレートを下げれば、エイリアスの影響が小さくなる。一般にエイリアスの影響はノイズとなり、復号時のエラーレートに影響する。

10

【0035】

(A - 4) ところで、M等帯域分割最大間引きフィルタバンクに基づくサブバンド符号化における符号のレートは、前述のように出力波形の標本化周波数の $1/M$ となる。各サブバンドの(ダウンサンプリング後の)周波数成分が $-f_s/2M$ から $f_s/2M$ (ここでは f_s は出力の標本化周波数)の範囲の成分であると考えた場合、周波数 $f_s/4M$ を搬送波の周波数とするデジタル変調結果を直接サブバンドの成分とすることができる。 $f_s/4M$ はサブバンド符号化および復号における各バンドパスフィルタの中心周波数に対応し、先述の、当該バンドの中心周波数を搬送波の周波数とするデジタル変調は、 $1/4M$ を搬送波の周波数とするデジタル変調結果を当該サブバンドの符号値(サブバンド符号ベクトルにおける当該サブバンドに対応する要素)として、それを復号した結果に対応する。すなわち、サブバンド符号の符号レートの $1/4$ の周波数を搬送波の周波数とするデジタル変調結果をサブバンド符号とすれば良い。例えば、信号点が振幅1、位相0に対応する変調結果は、サブバンド符号の当該サブバンドの要素の時系列データ上では、1、0、-1、0の繰り返しになる。また振幅1、位相 $\pi/4$ の場合は、 $1/\sqrt{2}$ 、 $1/\sqrt{2}$ 、 $-1/\sqrt{2}$ 、 $-1/\sqrt{2}$ の繰り返しになる。すなわち、サブバンドの中心周波数を搬送波の周波数とするデジタル変調は、サブバンド上で低レートかつ容易に実現できる。

20

30

【0036】

(A - 5) このようなデジタル変調結果を符号ベクトルの要素とするサブバンド符号を復号すると、復号処理におけるアップサンプリングとバンドパスフィルタによる帯域制限により、当該サブバンドの中心周波数を搬送波の周波数とするデジタル変調結果と等価になる。したがって、復号結果に対して、その中心周波数のコサイン関数およびサイン関数との相関を求めることでデジタル変調された結果を復調できる。なお復調の処理レートを下げるために、先の最終的な(復号された)波形に対してサブバンド符号化の際と同様の帯域制限、ダウンサンプリングを行い、その結果に対して周波数 $f_s/4M$ のコサイン関数、サイン関数との相関を求める方法とすることも可能である。

【0037】

40

(A - 6) 搬送波の周波数として取り得る値は、先述のようにサブバンド符号化におけるフィルタバンクの中心周波数となるので、デジタル変調方式としては、基本的に搬送波の周波数が一定である方式を用いる必要があり、そのような方式として、例えば、ASK(振幅偏移変調)やPSK(位相偏移変調)、QAM(直角位相振幅変調)がある。また、複数のサブバンドをデジタル変調に同時に用いる場合は、FSK(周波数偏移変調)を用いることもできる。

【0038】

(A - 7) なお、帯域制限された波形に対するサブバンド符号ベクトルはスパースになるので、変調部のバンド構成と、サブバンド符号化のサブバンド構成(フィルタバンク構成)を揃えなくても、処理効率の面ではまだ有利である。例えば隣接する2つのサブバン

50

ドの周波数帯域で行うような、より広帯域なデジタル変調も考えられる。このように、隣接する複数のサブバンドをまとめて用いると、変調部3のバンド数は減るものの、エイリアスの影響をほとんど受けない帯域を作ることができる。例えば、一番低域側のサブバンドとその次のサブバンドを2つを組として使う場合、周波数 $f_s / 4M$ から $3f_s / 4M$ の範囲はエイリアスの影響をほとんど受けない(2つのサブバンド間でエイリアスを打ち消し合う)形にできる。この場合のデータレートは時間波形領域で同様の処理を行う場合の $2/M$ となる(サブバンド符号のM次元のベクトルのうち、2個の要素を除き、値が近似的に0になる)。対象とする音声波形合成装置においては、合成部でサブバンド符号の復号処理がそもそも必須であることから、デジタルデータ重畳方式間の処理量比較においてはこの処理を考慮に入れる必要がない。データレートが低くなるメリットがあることから、サブバンド符号上での処理は、依然有利である。

10

【0039】

(B) 音声波形合成装置10において、デジタルデータを重畳しても合成音声は聴覚上、自然なものとして得られるようにする観点等から、以下の(B-1)~(B-4)のようにしてもよい。

【0040】

(B-1) 合成音声にデジタルデータを重畳する場合、その音響的な影響を考慮する必要がある。搬送波の成分による聴感上の影響を抑えるためには、変調結果が定常的(ここでは、周期的な特徴も含む)にならないようにする必要がある。このためには例えば、ある程度長い周期を持った疑似乱数系列との排他的論理和(XOR)を求めた結果を用いるなどのスクランブル処理を導入すればよい。あるいは、変調方式に $\pi/4$ シフトQPSKのような、シンボルごとに位相が必ず変わる方式を用いても良い。またそのような変調結果は、聴感上は雑音的な音響信号となるので、変調結果で合成音声の雑音部分を置き換える形にすることで、聴感上の影響を抑えることができる。また、一般に音声の高域については、声帯振動周期の揺らぎによりその調波成分が崩れ、そのスペクトル包絡特性が等しければ、雑音と置き換えても聴感上の差異は小さいので、高域のサブバンドにデジタルデータを埋め込むことは容易である。聴感上の影響を抑えるためには、当該サブバンドの短時間(例えば数十ミリ秒)のパワーが、音声のそれと等しくなるようにすれば良い。この場合、その絶対的な振幅が復調において問題にならないようにする必要がある。ただし、QAM等でも適当な時間範囲内に大きい振幅のシンボルが含まれるような制限を設けて復調可能にすれば用いることができる。

20

30

【0041】

(B-2) また、長時間連続して特定のバンドにデジタルデータを重畳すると、その雑音上の波形成分が聴感上問題となる場合も考えられるので、例えば1シンボルの時間的な長さを5msとしてそれを1タイムスロットとし、全体のうちの例えば1割や2割といった所定の基準により、一部のタイムスロットのみデジタルデータを重畳する方法を用いても良い。この場合も、デジタルデータを重畳するタイムスロットを時間的に等間隔に重畳すると、特にタイムスロットの単位を短くした場合に、タイムスロットの配置周期に対応する周波数成分が聴こえてしまう可能性があるため、疑似乱数等により、時間的にランダムになるように重畳することで聴感的な影響を緩和できる。これらの処理は音声波形生成装置10における変調部3でバンド毎に独立に制御できるので、デジタルデータを重畳するバンドも(疑似的に)ランダムとなるような配置にすることで、より影響を抑えることができる。

40

【0042】

(B-3) また、聴感上の影響が少ない高域バンドに、どのバンドにデジタルデータを埋め込んだかの情報を格納したかを情報として格納してもよい。この情報を使うことで、低域であっても聴感上の影響が小さいサブバンド(例えば非周期性が高く雑音的な音響的特徴となっているサブバンド)にのみ選択的にデジタルデータを埋め込むことができる。この方法により、そのような処理を行わない場合と比較し、聴感上の影響を抑えつつ、より低域のサブバンドにデジタルデータを埋め込むことができる。この場合、高域サブバン

50

ドとして中心周波数が一番高いサブバンドを用いても良いし、一般的な音響システムの高域遮断特性等により合成音声波形の高域が音響的に遮断されてしまうことも考慮し、聴感上の影響が小さい範囲で、多少低い周波数のサブバンドを用いても良い。

【 0 0 4 3 】

(B - 4) さらに、例えば高域バンドに合成音声の非周期性情報やスペクトル包絡特性に関する情報をデジタルデータとして重畳し、そこから定まる当該バンドの雑音源のパワーの情報を、デジタルデータが埋め込まれているかどうかの判定で用いる構成としても良い。すなわち、スペクトル包絡特性および非周期性情報に関して、所定の条件に該当する場合には所定の(低域)バンドにデジタルデータが重畳される、という規則を予め設定しておいたうえで、ほぼ常に聴感上の影響が小さい所定の(主に高域の)バンドに非周期性情報や、スペクトル包絡特性の情報を重畳するようにしてもよい。合成音声のスペクトル包絡特性は、ケプストラム係数でも良いが、音声合成に用いたものと同じ音声合成システムが使える場合は、最終的に非周期性情報やスペクトル包絡特性を得るための、音声合成の元のテキストや、音声合成に用いた音声合成記号列等の間接的な情報でも良い。

10

【 0 0 4 4 】

(C) また、本発明により、例えば、テキスト読み上げの合成音声に対して、そのテキスト自体をデジタルデータとして合成音声の音響信号に重畳しておけば、例えば音声認識技術によるものよりも正確に、その合成音声の内容を、その音響信号だけから表示するシステムを構築することができる。この際、このテキスト情報を、先に述べたような、どのタイムスロットにテキスト以外のデジタルデータを格納したのかの判定に使うことも可能である。

20

【 0 0 4 5 】

(D) なお、上記実施例では、周期性成分の合成に正弦波を用いたが、この方法に限定されない。音声に重畳するデジタルデータのデジタル変調結果の振幅を決めるために必要な、音声のスペクトル包絡特性が得られれば、第1のサブバンド符号ベクトルの生成方法について制限はなく、例えば、予め、多数の音声波形素片をサブバンド符号化したベクトルの形で蓄積しており、そのベクトルを、入力された音響パラメータの特徴を持つように、選択・接続する方法で、第1のサブバンド符号ベクトルを構成する、周期性成分のサブバンド符号ベクトルを生成しても良い。実施例に挙げた方法に限らないのは、雑音成分の生成においても同様である。

30

【 0 0 4 6 】

(E) また、上記実施例では、音声波形合成装置の入力として、スペクトル包絡特性に加えて基本周波数と非周期性情報を用いたが、例えば、ささやき声のように周期性を持たない音声のみを波形合成の対象とするのであれば、周期成分合成部そのものが不要となり、その音声のスペクトル包絡特性は、非周期性成分のスペクトル包絡特性と同じになる。すなわち、基本周波数と非周期性情報は不要となる。このような、周期性を持たない音声のみを対象とする波形合成装置に対しても本発明の手法を適用することができる。

【 0 0 4 7 】

(F) 図4は、一般的なコンピュータ装置70におけるハードウェア構成を示す図であり、以上説明してきた各実施形態の音声波形合成装置10は、このような構成を有する1つ以上のコンピュータ装置70として実現可能である。コンピュータ装置70は、所定命令を実行するCPU(中央演算装置)71、CPU71の実行命令の一部又は全部をCPU71に代わって又はCPU71と連携して実行する専用プロセッサ72(GPU(グラフィック演算装置)や深層学習専用プロセッサ等)、CPU71や専用プロセッサ72にワークエリアを提供する主記憶装置としてのRAM73、補助記憶装置としてのROM74、通信インタフェース75、ディスプレイ76、マウス、キーボード、タッチパネル等によりユーザ入力を受け付ける入力インタフェース77と、これらの間でデータを授受するためのバスBSと、を備える。

40

【 0 0 4 8 】

音声波形合成装置10の各部は、各部の機能に対応する所定のプログラムをROM74から読み込んで実行するCPU71及び/又は専用プロセッサ72によって実現することができる

50

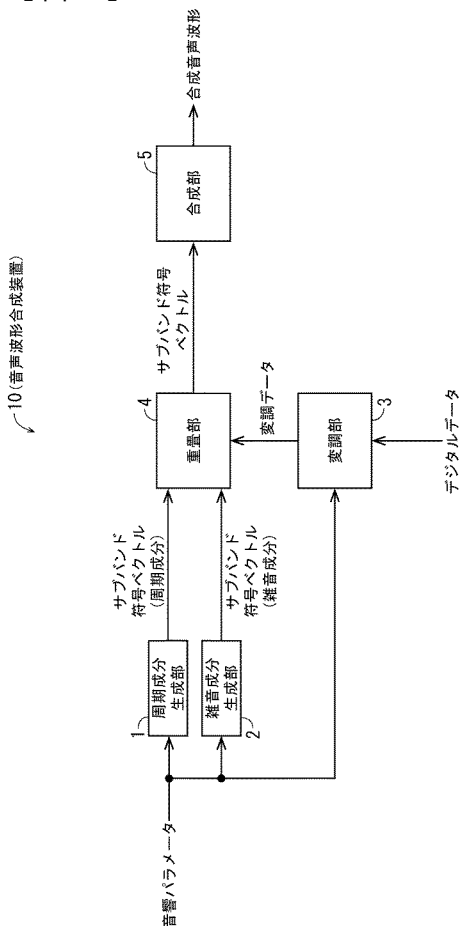
。また、音声波形合成装置10による音声波形合成方法は、当該方法に対応する所定のプログラムをROM74から読み込んで実行するCPU71及び/又は専用プロセッサ72によって実施することができる。

【符号の説明】

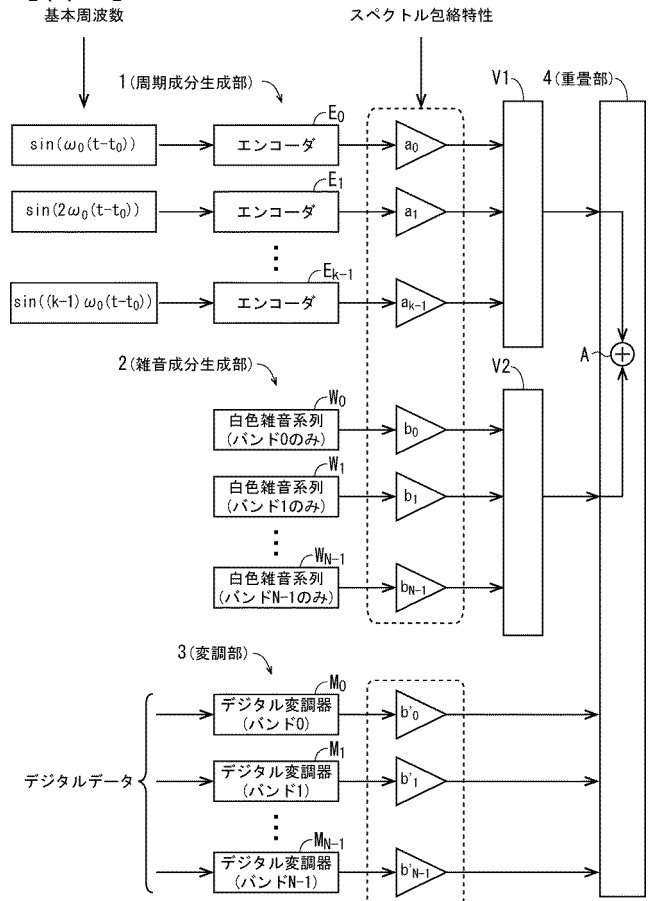
【0049】

10 音声波形合成装置、1 周期成分生成部、2 雑音成分生成部、3 変調部、4 重畳部、5 合成部

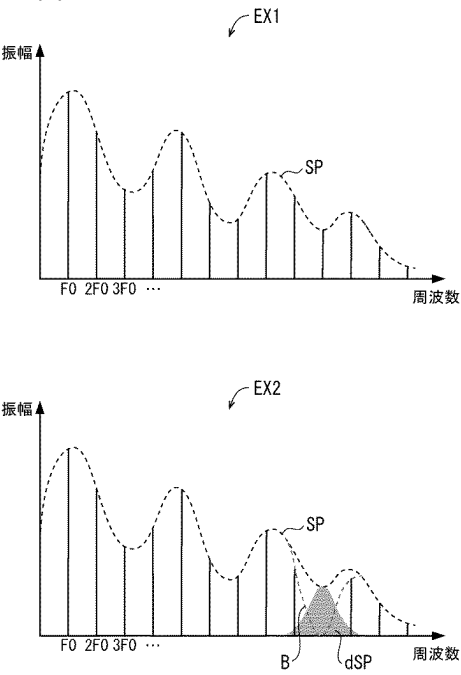
【図1】



【図2】



【図 3】



【図 4】

