(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7636088号
(P7636088)

(45)発行日　令和7年2月26日(2025.2.26)

(24)登録日　令和7年2月17日(2025.2.17)

(51)国際特許分類

		FI		
Ｇ１０Ｌ	*21/0264(2013.01)*	Ｇ１０Ｌ	21/0264	Ｃ
Ｇ１０Ｌ	*21/0232(2013.01)*	Ｇ１０Ｌ	21/0232	

請求項の数　15　（全37頁）

(21)出願番号　特願2023-527431(P2023-527431)
(86)(22)出願日　令和4年1月26日(2022.1.26)
(65)公表番号　特表2023-548707(P2023-548707A)
(43)公表日　令和5年11月20日(2023.11.20)
(86)国際出願番号　PCT/CN2022/074003
(87)国際公開番号　WO2022/166710
(87)国際公開日　令和4年8月11日(2022.8.11)
審査請求日　令和5年5月8日(2023.5.8)
(31)優先権主張番号　202110181389.4
(32)優先日　令和3年2月8日(2021.2.8)
(33)優先権主張国・地域又は機関
中国(CN)

(73)特許権者　517392436
▲騰▼▲訊▼科技（深▲セン▼）有限公司
ＴＥＮＣＥＮＴ ＴＥＣＨＮＯＬＯＧＹ（ＳＨＥＮＺＨＥＮ） ＣＯＭＰＡＮＹ ＬＩＭＩＴＥＤ
中華人民共和国５１８０５７▲広▼▲東▼省深▲セン▼市南山区高新区科技中一路 ▲騰▼▲訊▼大厦３５▲層▼
３５／Ｆ，Ｔｅｎｃｅｎｔ Ｂｕｉｌｄｉｎｇ，Ｋｅｊｉｚｈｏｎｇｙｉ Ｒｏａｄ，Ｍｉｄｗｅｓｔ Ｄｉｓｔｒｉｃｔ ｏｆ Ｈｉ－ｔｅｃｈ Ｐａｒｋ，Ｎａｎｓｈａｎ Ｄｉｓｔｒｉｃｔ，Ｓｈｅｎｚｈｅｎ，Ｇｕａｎｇｄｏｎｇ ５１８

最終頁に続く

(54)【発明の名称】　音声強調方法、装置、機器及びコンピュータプログラム

(57)【特許請求の範囲】

【請求項１】

コンピュータ機器によって実行される、音声強調方法であって、

目標音声フレームの対応する複素スペクトルに基づいて前記目標音声フレームに対してプリエンファシス処理を行い、第１複素スペクトルを得るステップと、

前記第１複素スペクトルに基づいて前記目標音声フレームに対して音声分解を行い、前記目標音声フレームの対応する声門パラメータ、ゲイン及び励起信号を得るステップと、

前記声門パラメータ、前記ゲイン及び前記励起信号に基づいて合成処理を行い、前記目標音声フレームの対応する強調音声信号を得るステップと

を含む、音声強調方法。

【請求項２】

目標音声フレームの対応する複素スペクトルに基づいて前記目標音声フレームに対してプリエンファシス処理を行い、第１複素スペクトルを得る前記ステップは、

前記目標音声フレームの対応する複素スペクトルを第１ニューラルネットワークに入力するステップであって、前記第１ニューラルネットワークはサンプル音声フレームの対応する複素スペクトルと前記サンプル音声フレームにおける元の音声信号の対応する複素スペクトルとに基づいてトレーニングを行って得られ、前記サンプル音声フレームは、前記元の音声信号とノイズ信号とを組み合わせることにより得られる、ステップと、

前記第１ニューラルネットワークによって、前記目標音声フレームの対応する複素スペクトルに基づいて前記第１複素スペクトルを出力するステップと

を含む、請求項１に記載の方法。
【請求項３】
前記第１ニューラルネットワークは複素畳み込み層、ゲート付き回帰型ユニット層及び全結合層を含み、
前記第１ニューラルネットワークによって、前記目標音声フレームの対応する複素スペクトルに基づいて前記第１複素スペクトルを出力する前記ステップは、
前記複素畳み込み層によって前記目標音声フレームに対応する複素スペクトルにおける実部及び虚部に基づいて複素畳み込み処理を行うステップと、
前記ゲート付き回帰型ユニット層によって前記複素畳み込み層の出力に対して変換処理を行うステップと、
前記全結合層によって前記ゲート付き回帰型ユニットの出力に対して全結合処理を行い、前記第１複素スペクトルを出力するステップと
を含む、請求項２に記載の方法。
【請求項４】
前記第１複素スペクトルに基づいて前記目標音声フレームに対して音声分解を行い、前記目標音声フレームの対応する声門パラメータ、ゲイン及び励起信号を得る前記ステップは、
前記第１複素スペクトルに基づいて前記目標音声フレームに対して声門パラメータ予測を行い、前記目標音声フレームの対応する声門パラメータを得るステップと、
前記第１複素スペクトルに基づいて前記目標音声フレームに対して励起信号予測を行い、前記目標音声フレームの対応する励起信号を得るステップと、
前記目標音声フレームの前の履歴音声フレームの対応するゲインに基づいて前記目標音声フレームに対してゲイン予測を行い、前記目標音声フレームの対応するゲインを得るステップと
を含む、請求項１に記載の方法。
【請求項５】
前記第１複素スペクトルに基づいて前記目標音声フレームに対して声門パラメータ予測を行い、前記目標音声フレームの対応する声門パラメータを得る前記ステップは、
前記第１複素スペクトルを第２ニューラルネットワークに入力するステップであって、前記第２ニューラルネットワークはサンプル音声フレームの対応する複素スペクトルと前記サンプル音声フレームの対応する声門パラメータとに基づいてトレーニングを行って得られるものである、ステップと、
前記第２ニューラルネットワークによって、前記第１複素スペクトルに基づいて前記目標音声フレームの対応する声門パラメータを出力するステップと
を含む、請求項４に記載の方法。
【請求項６】
前記第１複素スペクトルに基づいて前記目標音声フレームに対して声門パラメータ予測を行い、前記目標音声フレームの対応する声門パラメータを得る前記ステップは、
前記第１複素スペクトルと前記目標音声フレームの前の履歴音声フレームの対応する声門パラメータとを第２ニューラルネットワークに入力するステップであって、前記第２ニューラルネットワークはサンプル音声フレームの対応する複素スペクトル、サンプル音声フレームの前の履歴音声フレームの対応する声門パラメータ及びサンプル音声フレームの対応する声門パラメータに基づいてトレーニングを行って得られるものである、ステップと、
前記第２ニューラルネットワークによって、前記第１複素スペクトルと前記目標音声フレームの前の履歴音声フレームの対応する声門パラメータとに基づいて前記目標音声フレームの対応する声門パラメータを出力するステップと
を含む、請求項４に記載の方法。
【請求項７】
前記目標音声フレームの前の履歴音声フレームの対応するゲインに基づいて前記目標音

声フレームに対してゲイン予測を行い、前記目標音声フレームの対応するゲインを得る前記ステップは、
前記目標音声フレームの前の履歴音声フレームの対応するゲインを第３ニューラルネットワークに入力するステップであって、前記第３ニューラルネットワークはサンプル音声フレームの前の履歴音声フレームの対応するゲインと前記サンプル音声フレームの対応するゲインとに基づいてトレーニングを行って得られるものである、ステップと、
前記第３ニューラルネットワークによって、前記目標音声フレームの前の履歴音声フレームの対応するゲインに基づいて前記目標音声フレームの対応するゲインを出力するステップと
を含む、請求項４に記載の方法。
【請求項８】
前記第１複素スペクトルに基づいて前記目標音声フレームに対して励起信号予測を行い、前記目標音声フレームの対応する励起信号を得る前記ステップは、
前記第１複素スペクトルを第４ニューラルネットワークに入力するステップであって、前記第４ニューラルネットワークはサンプル音声フレームの対応する複素スペクトルと前記サンプル音声フレームに対応する励起信号の周波数領域表現とに基づいてトレーニングを行って得られるものである、ステップと、
前記第４ニューラルネットワークによって、前記第１複素スペクトルに基づいて前記目標音声フレームに対応する励起信号の周波数領域表現を出力するステップと
を含む、請求項４に記載の方法。
【請求項９】
前記声門パラメータ、前記ゲイン及び前記励起信号に基づいて合成処理を行い、前記目標音声フレームの対応する強調音声信号を得る前記ステップは、
声門フィルターにより前記目標音声フレームの対応する励起信号に対してフィルタリングを行い、フィルタリング出力信号を得るステップであって、前記声門フィルターは前記目標音声フレームの対応する声門パラメータに基づいて構築されるものである、ステップと、
前記目標音声フレームの対応するゲインに応じて前記フィルタリング出力信号に対して増幅処理を行い、前記目標音声フレームの対応する強調音声信号を得るステップと
を含む、請求項４に記載の方法。
【請求項１０】
前記第１複素スペクトルに基づいて前記目標音声フレームに対して音声分解を行い、前記目標音声フレームの対応する声門パラメータ、ゲイン及び励起信号を得る前記ステップは、
前記第１複素スペクトルに基づいてパワースペクトルを計算して取得するステップと、
前記パワースペクトルに基づいて自己相関係数を計算して取得するステップと、
前記自己相関係数に基づいて前記声門パラメータを計算して取得するステップと、
前記声門パラメータと前記自己相関係数とに基づいて前記ゲインを計算して取得するステップと、
前記ゲインと声門フィルターのパワースペクトルとに基づいて前記励起信号のパワースペクトルを計算して取得するステップであって、前記声門フィルターは前記声門パラメータに基づいて構築されるフィルターである、ステップと
を含む、請求項１に記載の方法。
【請求項１１】
前記声門パラメータ、前記ゲイン及び前記励起信号に基づいて合成処理を行い、前記目標音声フレームの対応する強調音声信号を得る前記ステップは、
前記声門フィルターのパワースペクトルと前記励起信号のパワースペクトルとに基づいて第１振幅スペクトルを生成するステップと、
前記ゲインに応じて前記第１振幅スペクトルに対して増幅処理を行い、第２振幅スペクトルを得るステップと、

前記第２振幅スペクトルと前記第１複素スペクトル中から抽出された位相スペクトルとに基づいて、前記目標音声フレームの対応する強調音声信号を決定するステップとを含む、請求項１０に記載の方法。

【請求項１２】

前記第２振幅スペクトルと前記第１複素スペクトル中から抽出された位相スペクトルとに基づいて、前記目標音声フレームの対応する強調音声信号を決定する前記ステップは、

前記第２振幅スペクトルと前記第１複素スペクトル中から抽出された位相スペクトルとを組み合わせ、第２複素スペクトルを得るステップと、

前記第２複素スペクトルを時間領域に変換し、前記目標音声フレームに対応する強調音声信号の時間領域信号を得るステップと

を含む、請求項１１に記載の方法。

【請求項１３】

音声強調装置であって、

目標音声フレームの複素スペクトルに基づいて前記目標音声フレームに対してプリエンファシス処理を行い、第１複素スペクトルを得ることに用いられるプリエンファシスモジュールと、

前記第１複素スペクトルに基づいて前記目標音声フレームに対して音声分解を行い、前記目標音声フレームの対応する声門パラメータ、ゲイン及び励起信号を得ることに用いられる音声分解モジュールと、

前記声門パラメータ、前記ゲイン及び前記励起信号に基づいて合成処理を行い、前記目標音声フレームの対応する強調音声信号を得ることに用いられる合成処理モジュールと

を含む、音声強調装置。

【請求項１４】

電子機器であって、

プロセッサと、

メモリであって、前記メモリ上にコンピュータ可読指令が記憶され、前記コンピュータ可読指令が前記プロセッサによって実行されるときに、請求項１～１２のいずれか一項に記載の方法を実現するメモリと

を含む、電子機器。

【請求項１５】

コンピュータプログラムであって、プロセッサによって実行されるときに、請求項１～１２のいずれか一項に記載の方法を実現する、コンピュータプログラム。

【発明の詳細な説明】

【技術分野】

【０００１】

本願は音声処理の技術分野に関し、具体的に言えば、音声強調方法、装置、機器及び記憶媒体に関する。

【０００２】

本願は２０２１年２月８日に中国特許庁に提出された、出願番号が第２０２１１０１８１３８９．４号、発明の名称が「音声強調方法、装置、機器及び記憶媒体」である中国特許出願の優先権を主張し、その全内容は引用により本願に組み込まれている。

【背景技術】

【０００３】

音声通信の便利性及び適時性により、音声通信の応用はますます幅広くなっており、たとえば、クラウド会議の会議参加者の間で音声信号が伝送される。ただし、音声通信においては、音声信号中にはノイズが混入される可能性があり、音声信号中に混入されるノイズが通信品質の劣化を招き、ユーザーの聴覚的体験に極めて大きな影響を与えることがある。従って、如何に音声に対して強調処理を行うことでノイズを除去するかは従来技術において早急に解決する技術的課題である。

【発明の概要】

【発明が解決しようとする課題】
【０００４】
本願の実施例は音声強調方法、装置、機器及び記憶媒体を提供することで、音声強調を実現し、音声信号の品質を向上させる。
【０００５】
本願のその他特性及び利点は以下の詳細な記述により明らかになるか、又は部分的に本願の実践により把握されて得られる。
【課題を解決するための手段】
【０００６】
本願の実施例の一態様によれば、音声強調方法を提供し、目標音声フレームの対応する複素スペクトルに基づいて前記目標音声フレームに対してプリエンファシス処理を行い、第１複素スペクトルを得るステップと、前記第１複素スペクトルに基づいて前記目標音声フレームに対して音声分解を行い、前記目標音声フレームの対応する声門パラメータ、ゲイン及び励起信号を得るステップと、前記声門パラメータ、前記ゲイン及び前記励起信号に基づいて合成処理を行い、前記目標音声フレームの対応する強調音声信号を得るステップとを含む。
【０００７】
本願の実施例の別の一態様によれば、音声強調装置を提供し、目標音声フレームの複素スペクトルに基づいて前記目標音声フレームに対してプリエンファシス処理を行い、第１複素スペクトルを得ることに用いられるプリエンファシスモジュールと、前記第１複素スペクトルに基づいて前記目標音声フレームに対して音声分解を行い、前記目標音声フレームの対応する声門パラメータ、ゲイン及び励起信号を得ることに用いられる音声分解モジュールと、前記声門パラメータ、前記ゲイン及び前記励起信号に基づいて合成処理を行い、前記目標音声フレームの対応する強調音声信号を得ることに用いられる合成処理モジュールとを含む。
【０００８】
本願の実施例の別の一態様によれば、電子機器を提供し、プロセッサと、メモリであって、前記メモリ上にコンピュータ可読指令が記憶され、前記コンピュータ可読指令が前記プロセッサによって実行されるときに、上記に記載の音声強調方法を実現するメモリとを含む。
【０００９】
本願の実施例の別の一態様によれば、コンピュータ可読記憶媒体を提供し、その上にコンピュータ可読指令が記憶され、前記コンピュータ可読指令がプロセッサによって実行されるときに、上記に記載の音声強調方法を実現する。
【発明の効果】
【００１０】
本願の解決手段においては、まず目標音声フレームに対してプリエンファシスを行って第１複素スペクトルを得て、次に第１複素スペクトルを基礎として目標音声フレームに対して音声分解と合成を行い、２段階に分けて目標音声フレームに対して強調を行うことを実現するため、音声強調効果を効果的に保証することができる。そして、目標音声フレームに対してプリエンファシスを行って得られた第１複素スペクトルを基礎として、目標音声フレームに対して音声分解を行い、プリエンファシス前の目標音声フレームに比べて、第１複素スペクトルにおけるノイズの情報がより少なくなる。一方、音声分解過程において、ノイズが音声分解の正確性に影響を与えることがあり、従って、第１複素スペクトルを音声分解の基礎とすることで、音声分解の難度を低減させ、音声分解で得られた声門パラメータ、励起信号及びゲインの正確性を向上させ、さらに後続で取得された強調音声信号の正確性を保証することができる。そして、プリエンファシスで得られた第１複素スペクトル中には位相情報と振幅情報とが含まれ、該第１複素スペクトルにおける位相情報と振幅情報とを基礎として音声分解及び音声合成を行うことで、得られた目標音声フレームに対応する強調音声信号の振幅と位相の精度が保証されている。

【００１１】
理解すべきことは、以上の一般的な記述と後述の細部の記述は例示的で解釈的なものに過ぎず、本願を限定し得るものではないことである。
【００１２】
ここでの図面は、明細書に組み込まれ、且つ本明細書の一部を構成しており、本願にマッチングする実施例を示し、且つ明細書とともに本願の原理を解釈することに用いられる。明らかなように、以下の記述における図面は本願のいくつかの実施例に過ぎず、当業者にとって、創造的な労働を必要としない前提において、これらの図面に基づいてその他の図面を取得することもできる。図面において以下のとおりである。
【図面の簡単な説明】
【００１３】
【図１】１つの具体的な実施例に基づいて示されるＶｏＩＰシステムにおける音声通信リンクの模式図である。
【図２】音声信号が生じているデジタルモデルの模式図を示す。
【図３】１つの元の音声信号に基づいて励起信号と声門フィルターを分解する周波数応答の模式図を示す。
【図４】本願の一実施例に基づいて示される音声強調方法のフローチャートである。
【図５】１つの具体的な実施例に基づいて示される複素畳み込み層が複素数に対して畳み込み処理を行う模式図である。
【図６】１つの具体的な実施例に基づいて示される第１ニューラルネットワークの構造模式図である。
【図７】１つの具体的な実施例に基づいて示される第２ニューラルネットワークの模式図である。
【図８】別の一実施例に基づいて示される第２ニューラルネットワークの入力と出力の模式図である。
【図９】１つの具体的な実施例に基づいて示される第３ニューラルネットワークの模式図である。
【図１０】１つの具体的な実施例に基づいて示される第４ニューラルネットワークの模式図である。
【図１１】一実施例に基づいて示されるステップ４３０のフローチャートである。
【図１２】１つの具体的な実施例に基づいて示される音声強調方法のフローチャートである。
【図１３】一実施例に基づいて示されるステップ４２０のフローチャートである。
【図１４】別の一実施例に基づいて示されるステップ４３０のフローチャートである。
【図１５】別の１つの具体的な実施例に基づいて示される音声強調方法のフローチャートである。
【図１６】１つの具体的な実施例に基づいて示される短時間フーリエ変換における窓掛け・オーバーラップの模式図である。
【図１７】一実施例に基づいて示される音声強調装置のブロック図である。
【図１８】本願の実施例を実現するための電子機器に適するコンピュータシステムの構造模式図を示す。
【発明を実施するための形態】
【００１４】
これより、図面を参照しながら例示的な実施形態をより全面的に記述する。しかしながら、例示的な実施形態は複数種の形式で実施でき、且つここで述べられた例に限定されると理解すべきでない。逆に、これらの実施形態の提供により、本願はより全面的で完全になり、且つ例示的な実施形態の発想は当業者に全面的に伝達される。
【００１５】
この他、記述される特徴、構造又は特性は、任意の適切な方式で１つ又はより多くの実施例に組み込まれてもよい。以下の記述において、多くの具体的な細部を提供することで

本願の実施例に対する十分な理解を与える。しかしながら、当業者は、特定の細部のうちの１つ又はより多くがなかったとしても、又はその他の方法、エレメント、装置、ステップ等を採用したとしても本願の技術的手段を実践できることを認識することができる。その他の状況においては、公知の方法、装置、実現又は操作を詳細に示さない、又は記述しないことによって、本願の各態様を不明瞭にすることを回避する。

【００１６】

図面において示されるブロック図は、単なる機能エンティティであり、必ずしも物理的に独立したエンティティに対応するわけではない。すなわち、ソフトウェアの形式を採用することでこれらの機能エンティティを実現する、又は１つ又は複数のハードウェアモジュール又は集積回路においてこれらの機能エンティティを実現する、又は異なるネットワーク及び／又はプロセッサ装置及び／又はマイクロ制御器装置においてこれらの機能エンティティを実現することができる。

【００１７】

図面において示されるフローチャートは例示的な説明に過ぎず、必ずしもあらゆる内容と操作／ステップを含むわけではなく、必ずしも記述された順序で実行されるわけでもない。たとえば、ある操作／ステップはさらに分解でき、一方、ある操作／ステップは併せることができ、又は部分的に併せることができ、従って、実際に実行される順序は実際の状況に応じて変化する可能性がある。

【００１８】

説明する必要がある点として、本明細書中に言及される「複数」は２つ又は２つ以上を指す。「及び／又は」は関連対象の関連関係を記述し、３種の関係が存在できることを表し、たとえば、Ａ及び／又はＢは、Ａが単独で存在すること、ＡとＢが同時に存在すること、Ｂが単独で存在することの３種の状況を表すことができる。文字「／」は一般的に前後の関連対象が「又は」の関係であることを表す。

【００１９】

音声信号におけるノイズが、音声品質を極めて大きく低減させ、ユーザーの聴覚的体験に影響を与えることがあり、従って、音声信号の品質を向上させるために、音声信号に対して強調処理を行うことで、ノイズを最大限に除去し、信号における元の音声信号（すなわち、ノイズを含まない純粋な信号）を保留する必要がある。音声に対して強調処理を行うことを実現するために、本願の解決手段が提案されている。

【００２０】

本願の解決手段は、音声通話の応用シーンにおいて適用でき、たとえば、インスタントメッセージングアプリケーションを介して行われる音声通信、ゲームアプリケーションにおける音声通話である。具体的には、音声の送信端、音声の受信端、又は音声通信サービスを提供するサーバ端末で本願の解決手段に従って音声強調を行うことができる。

【００２１】

クラウド会議はオンライン業務実行における１つの重要な過程であり、クラウド会議において、クラウド会議の参加者の音収集装置が発言者の音声信号を収集した後に、収集された音声信号をその他の会議参加者に送信する必要がある。該過程に関わる音声信号は複数の参加者の間で伝送されて再生され、音声信号中に混入されたノイズ信号に対して処理を行われなければ、会議参加者の聴覚的体験に極めて大きな影響を与えることがある。このようなシーンにおいて、本願の解決手段を応用してクラウド会議中の音声信号に対して強調を行うことができ、これにより、会議参加者が聞き取っる音声信号は強調された後の音声信号とすることができ、音声信号の品質を向上させることができる。

【００２２】

クラウド会議は、クラウドコンピューティング技術に基づく高効率で、便利な、低コストの会議形式である。ユーザーはインターネットインターフェースを介して、簡単で使いやすい操作を行うだけで、迅速且つ高効率に世界的なチーム及び顧客と音声、データファイル及びビデオを同期して共有することができ、一方、会議中のデータの伝送、処理等の複雑な技術はクラウド会議サービス提供者がユーザーを補助することにより操作され得る。

【００２３】
現在、中国国内のクラウド会議は主にＳａａＳ（Ｓｏｆｔｗａｒｅ ａｓ ａ Ｓｅｒｖｉｃｅ、ソフトウェア・アズ・ア・サービス）モードを主体とするサービス内容に焦点を当てて、電話、ネットワーク、ビデオ等のサービス形式を含み、クラウドコンピューティングに基づくビデオ会議はクラウド会議と呼ばれる。クラウド会議の時代においては、データの伝送、処理、記憶はすべてビデオ会議提供者のコンピュータリソースにより処理され、ユーザーはさらに高価なハードウェアを購入したり煩雑なソフトウェアをインストールしたりする必要が全くなく、クライアント端末を開いて対応するインターフェースにアクセスするだけで、高効率な遠隔会議を行うことができる。
【００２４】
クラウド会議システムは、マルチサーバの動的クラスター配置をサポートし、且つ複数台の高性能サーバを提供し、会議の安定性、安全性、可用性を大幅に高める。近年、ビデオ会議はコミュニケーション効率を大幅に向上させ、コミュニケーションコストを連続的に低減させ、内部管理レベルのアップグレードをもたらすことができるため、多くのユーザーに人気があり、すでに政府、軍隊、交通、輸送、金融、オペレータ、教育、企業等の各分野に幅広く応用されている。
【００２５】
図１は、１つの具体的な実施例に基づいて示されるＶｏＩＰ（Ｖｏｉｃｅ ｏｖｅｒ Ｉｎｔｅｒｎｅｔ Ｐｒｏｔｏｃｏｌ、ネットワーク電話）システムにおける音声通信リンクの模式図である。図１に示すように、送信端１１０と受信端１２０のネットワーク接続に基づき、送信端１１０と受信端１２０は音声伝送を行うことができる。
【００２６】
図１に示すように、送信端１１０は収集モジュール１１１、前強調処理モジュール１１２及び符号化モジュール１１３を含み、ここで、収集モジュール１１１は、音声信号を収集することに用いられ、それは収集した音響信号をデジタル信号に変換することができ、前強調処理モジュール１１２は、収集された音声信号に対して強調を行うことで、収集された音声信号中のノイズを除去し、音声信号の品質を向上させることに用いられる。符号化モジュール１１３は、強調された後の音声信号に対して符号化を行うことで、音声信号の伝送過程中の干渉抵抗性を向上させることに用いられる。前強調処理モジュール１１２は、本願の方法に従って音声強調を行い、音声に対して強調を行った後、さらに符号化圧縮及び伝送を行うことができ、このように、受信端が受信した信号がノイズに影響されなくなることを保証できる。
【００２７】
受信端１２０は復号モジュール１２１、後強調モジュール１２２及び再生モジュール１２３を含む。復号モジュール１２１は受信した符号化音声信号に対して復号を行い、復号後の音声信号を得ることに用いられ、後強調モジュール１２２は復号後の音声信号に対して強調処理を行うことに用いられ、再生モジュール１２３は強調処理後の音声信号を再生することに用いられる。後強調モジュール１２２は本願の方法に従って音声強調を行うこともできる。いくつかの実施例では、受信端１２０はさらに音響効果調節モジュールを含んでもよく、該音響効果調節モジュールは強調された後の音声信号に対して音響効果調節を行うことに用いられる。
【００２８】
具体的な実施例において、受信端１２０のみ、又は送信端１１０のみで本願の方法に従って音声強調を行うことができ、もちろん、さらに送信端１１０と受信端１２０の両方で本願の方法に従って音声強調を行うこともできる。
【００２９】
いくつかの応用シーンにおいて、ＶｏＩＰシステムにおける端末機器はＶｏＩＰ通信をサポートできる以外に、さらにその他のサードパーティプロトコル、たとえば従来のＰＳＴＮ（Ｐｕｂｌｉｃ Ｓｗｉｔｃｈｅｄ Ｔｅｌｅｐｈｏｎｅ Ｎｅｔｗｏｒｋ、公共交換電話網）回路ドメイン電話をサポートすることもできる。一方、従来のＰＳＴＮサービス

は音声強調を行うことができず、このようなシーンにおいては、受信端としての端末において本願の方法に従って音声強調を行うことができる。

【００３０】

本願の解決手段に対して具体的な説明を行う前に、音声信号が生じるということについて説明を行う必要がある。音声信号は、人体の発音器官の脳制御における生理的運動によって生じるものであり、すなわち、気管のところで一定のエネルギーのノイズのような衝撃信号（励起信号に相当）が生じ、衝撃信号が人間の声帯（声帯が声門フィルターに相当）に衝撃を与え、略周期的な開閉が生じ、口腔を通じて増幅した後に、音を発する（音声信号を出力）。

【００３１】

図２は、音声信号が生じているデジタルモデルの模式図を示しており、該デジタルモデルにより音声信号が生じる過程を記述することができる。図２に示すように、励起信号は声門フィルターに衝撃を与えた後、さらにゲイン制御を行って、その後音声信号を出力し、ここで、声門フィルターは声門パラメータにより限定される。該過程は下式で表すことができる。

$$x(n)=G\cdot r(n)\cdot ar(n)\quad(式１)$$

ここで、ｘ（ｎ）は入力された音声信号を表し、Ｇはゲインを表し、線形予測ゲインと呼ばれることもでき、ｒ（ｎ）は励起信号を表し、ａｒ（ｎ）は声門フィルターを表す。

【００３２】

図３は、１つの元の音声信号に基づいて励起信号と声門フィルターを分解する周波数応答の模式図を示す。図３ａは該元の音声信号の周波数応答の模式図を示し、図３ｂは該元の音声信号に基づいて分解された声門フィルターの周波数応答の模式図を示し、図３ｃは該元の音声信号に基づいて分解された励起信号の周波数応答の模式図を示す。図３に示すように、該元の音声信号の周波数応答の模式図における波形部分は声門フィルターの周波数応答の模式図におけるピーク位置に対応し、励起信号は該元の音声信号に対してＬＰ（Ｌｉｎｅａｒ　Ｐｒｅｄｉｃｔｉｏｎ、線形予測）分析を行った後の残差信号に相当し、従って、その対応する周波数応答が比較的緩やかである。

【００３３】

上記からわかるように、１つの元の音声信号（すなわち、ノイズを含まない音声信号）に基づいて励起信号、声門フィルター及びゲインを分解することができ、分解された励起信号、声門フィルター及びゲインは該元の音声信号を表現することに用いられてもよく、ここで、声門フィルターは声門パラメータにより表現できる。逆に、１つの元の音声信号の対応する励起信号、声門フィルターを決定することに用いられる声門パラメータ及びゲインが知られていれば、対応する励起信号、声門フィルター及びゲインに基づいて該元の音声信号を再構成することができる。

【００３４】

本願の解決手段は、該原理に基づき、音声フレームの対応する声門パラメータ、励起信号及びゲインに基づいて該音声フレームにおける元の音声信号を再構成し、音声強調を実現することである。

【００３５】

以下、本願の実施例の技術的手段を詳細に述べる。

【００３６】

図４は、本願の一実施例に基づいて示される音声強調方法のフローチャートであり、該方法は処理能力を備えるコンピュータ機器により実行されてもよく、たとえば、端末、サーバ等であり、ここで具体的な限定を行わない。図４に示されるものを参照すると、該方法は少なくともステップ４１０～４３０を含み、以下のように詳細に説明される。

【００３７】

ステップ４１０：目標音声フレームの対応する複素スペクトルに基づいて前記目標音声フレームに対してプリエンファシス処理を行い、第１複素スペクトルを得る。

【００３８】

音声信号は緩やかでランダムに変化するのではなく経時的に変化するものであるが、短時間内で音声信号が強い相関を有する、すなわち、音声信号が短時間相関性を有する。従って、本願の解決手段において、音声フレームを単位として音声強調を行う。目標音声フレームとは現在の強調処理対象の音声フレームを指す。

【００３９】

目標音声フレームの対応する複素スペクトルは該目標音声フレームの時間領域信号に対して時間周波数変換を行うことにより取得することができ、時間周波数変換はたとえば短時間フーリエ変換（Ｓｈｏｒｔ－ｔｅｒｍ Ｆｏｕｒｉｅｒ ｔｒａｎｓｆｏｒｍ、ＳＴＦＴ）であってもよい。目標音声フレームの対応する複素スペクトルにおける実部の係数は該目標音声フレームの振幅情報を指示することに用いられ、虚部の係数は目標音声フレームの位相情報を指示することに用いられる。

【００４０】

目標音声フレームに対してプリエンファシスを行うことにより、目標音声フレームにおける一部のノイズを除去することができ、従って、目標音声フレームの対応する複素スペクトルに比べて、プリエンファシスで得られた第１複素スペクトルにおけるノイズ含有量がより少ない。

【００４１】

本願のいくつかの実施例では、深層学習の方式を採用して目標音声フレームに対してプリエンファシスを行うことができる。１つのニューラルネットワークモデルをトレーニングすることにより、音声フレームの対応する複素スペクトルに基づいて音声フレームにおけるノイズの複素スペクトルを予測し、次に音声フレームの複素スペクトルと予測されたノイズの複素スペクトルとを減算し、第１複素スペクトルを得る。記述の便宜のために、音声フレームにおけるノイズの複素スペクトルを予測することに用いられる該ニューラルネットワークモデルをノイズ予測モデルと呼ぶ。トレーニング終了後に、該ノイズ予測モデルは入力された音声フレームの複素スペクトルに基づいて予測されたノイズの複素スペクトルを出力することができ、次に音声フレームの複素スペクトルとノイズの複素スペクトルとを減算すると、第１複素スペクトルを得られる。

【００４２】

本願のいくつかの実施例では、さらに１つのニューラルネットワークモデルをトレーニングすることで、音声フレームの複素スペクトルに基づいて強調された後の該音声フレームの第１複素スペクトルを予測することができる。記述の便宜のために、強調された後の複素スペクトルを予測することに用いられる該ニューラルネットワークモデルを強調複素スペクトル予測モデルと呼ぶ。トレーニング過程において、サンプル音声フレームの複素スペクトルを該強調複素スペクトル予測モデル中に入力し、該強調複素スペクトル予測モデルによって強調された後の複素スペクトルを予測し、且つ予測された強調された後の複素スペクトルと該サンプル音声フレームのラベル情報とに基づいて強調複素スペクトル予測モデルのパラメータを調整し、予測された強調された後の複素スペクトルとラベル情報が指示した複素スペクトルとの間の差異が所定の要件を満たすまで続ける。サンプル音声フレームのラベル情報はサンプル音声フレームにおける元の音声信号の複素スペクトルを指示することに用いられる。トレーニング終了後に、該強調複素スペクトル予測モデルは目標音声フレームの複素スペクトルに基づいて第１複素スペクトルを出力することができる。

【００４３】

ステップ４２０：前記第１複素スペクトルに基づいて前記目標音声フレームに対して音声分解を行い、前記目標音声フレームの対応する声門パラメータ、ゲイン及び励起信号を得る。

【００４４】

音声分解で得られた目標音声フレームの対応する声門パラメータ、対応するゲイン及び対応する励起信号は、図２に示される過程に従って目標音声フレームにおける元の音声信号を再構成することに用いられる。

【００４５】
上記の記述のように、１つの元の音声信号は、励起信号が声門フィルターに衝撃を与えてからゲイン制御を行うことにより得られるものである。該第１複素スペクトル中には目標音声フレームの元の音声信号の情報が含まれており、従って、該第１複素スペクトルに基づき線形予測分析を行い、目標音声フレームにおける元の音声信号を再構成することに用いられる声門パラメータ、励起信号及びゲインを逆方向に決定する。
【００４６】
声門パラメータとは、声門フィルターを構築することに用いられるパラメータを指し、声門パラメータが決定されると、声門フィルターが対応して決定され、声門フィルターはデジタルフィルターである。声門パラメータは線形予測符号化（Ｌｉｎｅａｒ　Ｐｒｅｄｉｃｔｉｏｎ　Ｃｏｅｆｆｉｃｉｅｎｔｓ、ＬＰＣ）係数であってもよく、さらに線スペクトル周波数（Ｌｉｎｅ　Ｓｐｅｃｔｒａｌ　Ｆｒｅｑｕｅｎｃｙ、ＬＳＦ）パラメータであってもよい。目標音声フレームに対応する声門パラメータの数量は声門フィルターの次数に関連しており、前記声門フィルターがＫ次フィルターである場合、前記声門パラメータはＫ次ＬＳＦパラメータ又はＫ次ＬＰＣ係数を含み、ここで、ＬＳＦパラメータとＬＰＣ係数との間が相互に転換することができる。
【００４７】
１つのｐ次の声門フィルターは、
$A_p(z) = 1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_p z^{-p}$　（式２）として表されてもよい。
ここで、a_1、a_2、…、a_pはＬＰＣ係数であり、ｐは声門フィルターの次数であり、ｚは声門フィルターの入力信号である。
【００４８】
式２を基礎として、
$P(z) = A_p(z) - z^{-(p+1)} A_p(z^{-1})$　（式３）
$Q(z) = A_p(z) + z^{-(p+1)} A_p(z^{-1})$　（式４）のように設定する場合、
以下［数１］　（式５）を得ることができる。
【００４９】
【数１】

$$A_p(z) = \frac{P(z) + Q(z)}{2}$$

【００５０】
物理的には、Ｐ（ｚ）とＱ（ｚ）は、それぞれ声門開放と声門閉鎖の周期的な変化規律を代表する。多項式Ｐ（ｚ）とＱ（ｚ）の根は複素平面上で交互に出現し、それは複素平面単位円上に分布する一連の角周波数であり、ＬＳＦパラメータはすなわちＰ（ｚ）とＱ（ｚ）の根の複素平面単位円上の対応する角周波数であり、第ｎフレームの音声フレームの対応するＬＳＦパラメータＬＳＦ（ｎ）はωｎとして表されてもよい。もちろん、第ｎフレームの音声フレームの対応するＬＳＦパラメータＬＳＦ（ｎ）はさらに該第ｎフレームの音声フレームに対応するＰ（ｚ）の根と対応するＱ（ｚ）根で直接的に示されることができる。
【００５１】
第ｎフレームの音声フレームに対応するＰ（ｚ）とＱ（ｚ）の複素平面での根をθ_nとして定義すると、第ｎフレームの音声フレームの対応するＬＳＦパラメータは、
以下［数２］　（式６）として表される。
【００５２】
【数２】

$$\omega_n = tan^{-1}\left(\frac{Real\{\theta_n\}}{Imag\{\theta_n\}}\right)$$

【００５３】
ここで、Ｒｅｌ｛θｎ｝は複素数θｎの実部を表し、Ｉｍａｇ｛θｎ｝は複素数θｎの虚部を表す。
【００５４】
本願のいくつかの実施例では、深層学習の方式を採用して音声分解を行うことができる。まず、それぞれ声門パラメータ予測を行うこと、励起信号予測を行うこと、及びゲイン予測を行うことに用いられるニューラルネットワークモデルをトレーニングすることができ、該３つのニューラルネットワークモデルが第１複素スペクトルに基づき目標音声フレームの対応する声門パラメータ、励起信号及びゲインをそれぞれ予測できるようにする。
【００５５】
本願のいくつかの実施例では、さらに線形予測分析の原理に従って、第１複素スペクトルに基づいて信号処理を行い、且つ目標音声フレームの対応する声門パラメータ、励起信号及びゲインを計算することができ、具体的な過程は下記の記述を参照する。
【００５６】
ステップ４３０：前記声門パラメータ、前記ゲイン及び前記励起信号に基づいて合成処理を行い、前記目標音声フレームの対応する強調音声信号を得る。
【００５７】
目標音声フレームの対応する声門パラメータが決定される場合に、その対応する声門フィルターは対応して決定される。それを基に、図２に示される元の音声信号の生成過程に基づいて、目標音声フレームの対応する励起信号が決定される声門フィルターに衝撃を与え、且つ目標音声フレームの対応するゲインに応じてフィルタリングで得られた信号に対してゲイン制御を行うことにより、元の音声信号の再構成を実現することができ、再構成で取得された信号はすなわち目標音声フレームの対応する強調音声信号である。
【００５８】
本願の解決手段において、まず、目標音声フレームに対してプリエンファシスを行って第１複素スペクトルを得て、次に第１複素スペクトルを基礎として目標音声フレームに対して音声分解と合成を行い、２段階に分けて目標音声フレームに対して強調を行うことを実現し、音声強調効果を効果的に保証することができる。そして、目標音声フレームに対してプリエンファシスを行って得られた第１複素スペクトルを基礎として、目標音声フレームに対して音声分解を行い、目標音声フレームがプリエンファシスされる前のスペクトルに比べて、第１複素スペクトルにおけるノイズの情報がより少なくなる。音声分解過程においては、ノイズが音声分解の正確性に影響を与えることがあり、従って、第１複素スペクトルを音声分解の基礎とすることで、音声分解の難度を低減させ、音声分解で得られた声門パラメータ、励起信号及びゲインの正確性を向上させ、さらに後続で取得された強調音声信号の正確性を保証することができる。プリエンファシスで得られた第１複素スペクトル中には位相情報と振幅情報が含まれ、該第１複素スペクトルにおける位相情報と振幅情報を基礎として音声分解及び音声合成を行うことで、得られた目標音声フレームに対応する強調音声信号の振幅と位相の精度が保証されている。
【００５９】
本願のいくつかの実施例では、ステップ４１０は、前記目標音声フレームの対応する複素スペクトルを第１ニューラルネットワークに入力するステップであって、前記第１ニューラルネットワークはサンプル音声フレームの対応する複素スペクトルと前記サンプル音声フレームにおける元の音声信号の対応する複素スペクトルとに基づいてトレーニングを行って得られるものである、ステップと、前記第１ニューラルネットワークによって、前記目標音声フレームの対応する複素スペクトルに基づいて前記第１複素スペクトルを出力するステップとを含む。

【００６０】
第１ニューラルネットワークは、長・短期記憶ニューラルネットワーク、畳み込みニューラルネットワーク、回帰型ニューラルネットワーク、全結合ニューラルネットワーク、ゲート付き回帰型ユニット等により構築されたモデルであってもよく、ここで具体的な限定を行わない。
【００６１】
本願のいくつかの実施例では、サンプル音声信号に対してフレーム分割を行うことにより、複数のサンプル音声フレームを得ることができる。ここで、サンプル音声信号は、知られている元の音声信号と知られているノイズ信号とを組み合わせることにより得ることができ、このように、元の音声信号が知られている場合に、対応してサンプル音声フレームにおける元の音声信号に対して時間周波数変換を行って、サンプル音声フレームにおける元の音声信号の対応する複素スペクトルを得ることができる。サンプル音声フレームの対応する複素スペクトルは、該サンプル音声フレームの時間領域信号に対して時間周波数変換を行うことにより得ることができる。
【００６２】
トレーニング過程において、サンプル音声フレームの対応する複素スペクトルを第１ニューラルネットワークに入力し、第１ニューラルネットワークによって、サンプル音声フレームの対応する複素スペクトルに基づいて予測を行い、予測された第１複素スペクトルを出力し、次に予測された第１複素スペクトルと該サンプル音声フレームにおける元の音声信号の対応する複素スペクトルとを比較し、両方の間の類似度が所定の要件を満たさなければ、第１ニューラルネットワークのパラメータを調整し、第１ニューラルネットワークが出力した予測された第１複素スペクトルと該サンプル音声フレームにおける元の音声信号の対応する複素スペクトルとの間の類似度が所定の要件を満たすまで続ける。ここで、該所定の要件は、予測された第１複素スペクトルと該サンプル音声フレームにおける元の音声信号の対応する複素スペクトルとの間の類似度が類似度閾値以上であることであってもよく、該類似度閾値はニーズに応じて設定を行うことができ、たとえば、１００％、９８％等である。上記のようなトレーニング過程により、該第１ニューラルネットワークは入力された複素スペクトルに基づいて第１複素スペクトルを予測する能力を学習することができる。
【００６３】
本願のいくつかの実施例では、前記第１ニューラルネットワークは複素畳み込み層、ゲート付き回帰型ユニット層及び全結合層を含む。上記した前記第１ニューラルネットワークによって、前記目標音声フレームの複素スペクトルに基づいて前記第１複素スペクトルを出力するステップは、さらに、前記複素畳み込み層によって前記目標音声フレームに対応する複素スペクトルにおける実部及び虚部に基づいて複素畳み込み処理を行うステップと、前記ゲート付き回帰型ユニット層によって前記複素畳み込み層の出力に対して変換処理を行うステップと、前記全結合層によって前記ゲート付き回帰型ユニットの出力に対して全結合処理を行い、前記第１複素スペクトルを出力するステップとを含む。
【００６４】
具体的な実施例において、第１ニューラルネットワークは１層又は複数層の複素畳み込み層を含んでもよく、同様に、ゲート付き回帰型ユニット層と全結合層も１層又は複数層であってもよく、具体的には、複素畳み込み層、ゲート付き回帰型ユニット層及び全結合層の数量は実際のニーズに応じて設定を行うことができる。
【００６５】
図５は、１つの具体的な実施例に基づいて示される複素畳み込み層が複素数に対して畳み込み処理を行う模式図であり、複素畳み込み層の入力複素数がＥ＋ｊＦであり、複素畳み込み層の加重がＡ＋ｊＢであると仮定する。図５に示すように、複素畳み込み層は２次元畳み込み層（Ｒｅａｌ＿ｃｏｎｖ、Ｉｍａｇ＿ｃｏｎｖ）、結合層（Ｃｏｎｃａｔ）及び活性化層（Ｌｅａｋｙ＿Ｒｅｌｕ）を含む。入力複素数中の実部Ｅと虚部Ｆとを２次元畳み込み層に入力した後に、該２次元畳み込み層は複素畳み込み層の加重に応じて畳み込

みを行い、それが畳み込み演算を行う過程は下式で示される。
（Ｅ＋ｊＦ）＊（Ａ＋ｊＢ）＝（Ｅ＊Ａ－Ｆ＊Ｂ）＋ｊ（Ｅ＊Ｂ＋Ｆ＊Ａ）　（式７）
Ｃ＝Ｅ＊Ａ－Ｆ＊Ｂ、Ｄ＝Ｅ＊Ｂ＋Ｆ＊Ａに設定する場合、上式７はさらに、
（Ｅ＋ｊＦ）＊（Ａ＋ｊＢ）＝Ｃ＋ｊＤ　（式８）に転換する。

【００６６】

図５に示すように、２次元畳み込み層が畳み込まれた後の実部と虚部を出力した後に、結合層によって実部と虚部とを結合し、結合結果を得て、次に、活性化層によって結合結果に対して活性化を行う。図５において、活性化層に使用された活性化関数がＬｅａｋｙ＿Ｒｅｌｕ活性化関数である。Ｌｅａｋｙ＿Ｒｅｌｕ活性化関数の表現式は、
ｆ（ｘ）＝ｍａｘ（ａｘ，ｘ）（ａが定数である）　（式９）である。

【００６７】

その他の実施例において、活性化層に使用された活性化関数はさらにその他の関数、たとえばｚＲｅｌｕ関数等であってもよく、ここで具体的な限定を行わない。

【００６８】

図６は、１つの具体的な実施例に基づいて示される第１ニューラルネットワークの構造模式図であり、図６に示すように、該第１ニューラルネットワークは、順にカスケード接続された６層の複素畳み込み層（Ｃｏｎｖ）、１層のゲート付き回帰型ユニット（Ｇａｔｅｄ　Ｒｅｃｕｒｒｅｎｔ　Ｕｎｉｔ、ＧＲＵ）層及び２層の全結合（Ｆｕｌｌ　Ｃｏｎｎｅｃｔｅｄ、ＦＣ）層を含む。目標音声フレームに対応する複素スペクトルＳ（ｎ）を該第１ニューラルネットワークに入力した後に、まず６層の複素畳み込み層によって順に複素畳み込み処理を行い、次にＧＲＵ層によって変換を行い、さらに２層のＦＣ層によって順次に全結合を行い、且つ最後の１層のＦＣ層によって第１複素スペクトルを出力する。ここで、各層の括弧内の数字は該層が出力した変数の次元を表す。図６に示される第１ニューラルネットワークにおいて、最後の１層のＦＣ層が出力した次元は３２２次元であり、１６１個のＳＴＦＴ係数中の実部と虚部を示すことに用いられる。

【００６９】

本願のいくつかの実施例では、ステップ４２０は、前記第１複素スペクトルに基づいて前記目標音声フレームに対して声門パラメータ予測を行い、前記目標音声フレームの対応する声門パラメータを得るステップと、前記第１複素スペクトルに基づいて前記目標音声フレームに対して励起信号予測を行い、前記目標音声フレームの対応する励起信号を得るステップと、前記目標音声フレームの前の履歴音声フレームの対応するゲインに基づいて前記目標音声フレームに対してゲイン予測を行い、前記目標音声フレームの対応するゲインを得るステップとを含む。

【００７０】

本願のいくつかの実施例では、声門パラメータ予測を行うことに用いられるニューラルネットワークモデル（第２ニューラルネットワークとして仮定）、ゲイン予測を行うニューラルネットワークモデル（第３ニューラルネットワークとして仮定）、及び励起信号予測を行うニューラルネットワークモデル（第４ニューラルネットワークとして仮定）をそれぞれトレーニングすることができる。ここで、該３種のニューラルネットワークモデルは長・短期記憶ニューラルネットワーク、畳み込みニューラルネットワーク、回帰型ニューラルネットワーク、全結合ニューラルネットワーク等により構築されたモデルであってもよく、ここで具体的な限定を行わない。

【００７１】

本願のいくつかの実施例では、上記した前記第１複素スペクトルに基づいて前記目標音声フレームに対して声門パラメータ予測を行い、前記目標音声フレームの対応する声門パラメータを得るステップは、さらに、前記第１複素スペクトルを第２ニューラルネットワークに入力するステップであって、前記第２ニューラルネットワークはサンプル音声フレームの対応する複素スペクトルと前記サンプル音声フレームの対応する声門パラメータとに基づいてトレーニングを行って得られるものである、ステップと、前記第２ニューラルネットワークによって、前記第１複素スペクトルに基づいて前記目標音声フレームの対応

する声門パラメータを出力するステップとを含む。

【００７２】

サンプル音声フレームの対応する複素スペクトルは、サンプル音声フレームの時間領域信号に対して時間周波数変換を行うことにより得られるものである。本願のいくつかの実施例では、サンプル音声信号に対してフレーム分割を行い、複数のサンプル音声フレームを得ることができる。サンプル音声信号は知られている元の音声信号と知られているノイズ信号とを組み合わせることにより得ることができる。このように、元の音声信号が知られている場合に、元の音声信号に対して線形予測分析を行うことによりサンプル音声フレームの対応する声門パラメータを得ることができ、換言すれば、サンプル音声フレームの対応する声門パラメータとはサンプル音声フレームにおける元の音声信号を再構成することに用いられる声門パラメータを指す。

【００７３】

トレーニング過程においては、サンプル音声フレームの複素スペクトルを第２ニューラルネットワークに入力した後に、第２ニューラルネットワークによって、サンプル音声フレームの複素スペクトルに基づいて声門パラメータ予測を行い、予測声門パラメータを出力し、次に、予測声門パラメータと該サンプル音声フレームの対応する声門パラメータとを比較し、両方が一致しなければ、第２ニューラルネットワークのパラメータを調整し、第２ニューラルネットワークがサンプル音声フレームの複素スペクトルに基づいて出力した予測声門パラメータが該サンプル音声フレームの対応する声門パラメータと一致するまで続ける。トレーニング終了後に、該第２ニューラルネットワークは、入力された音声フレームの複素スペクトルに基づいて該音声フレームにおける元の音声信号を再構成することに用いられる声門パラメータを正確に予測する能力を学習している。

【００７４】

図７は、１つの具体的な実施例に基づいて示される第２ニューラルネットワークの模式図である。図７に示すように、該第２ニューラルネットワークは、１層のＬＳＴＭ（Ｌｏｎｇ－Ｓｈｏｒｔ　Ｔｅｒｍ　Ｍｅｍｏｒｙ、長・短期記憶ネットワーク）層と３層のカスケード接続されたＦＣ（Ｆｕｌｌ　Ｃｏｎｎｅｃｔｅｄ、全結合）層とを含む。ここで、ＬＳＴＭ層は１つの隠れ層であり、それは２５６個のユニットを含み、ＬＳＴＭ層の入力は第ｎフレームの音声フレームの対応する第１複素スペクトルＳ’（ｎ）である。本実施例において、ＬＳＴＭ層の入力は３２１次元である。３層のカスケード接続されたＦＣ層において、前の２層のＦＣ層中には活性化関数σ（）が設定され、設定された活性化関数は第２ニューラルネットワークの非線形発現能力を増加することに用いられ、最後の１層のＦＣ層中には活性化関数が設定されず、該最後の１層のＦＣ層は分類器として分類出力を行う。図７に示すように、入力から出力への方向に沿って、３層のＦＣ層中にはそれぞれ５１２、５１２、１６個のユニットが含まれ、最後の１層のＦＣ層の出力は該第ｎフレームの音声フレームに対応する１６次元の線スペクトル周波数係数ＬＳＦ（ｎ）、すなわち１６次線スペクトル周波数パラメータである。

【００７５】

本願のいくつかの実施例では、音声フレームの間に相関性があり、隣接する２つの音声フレームの間の周波数領域特徴の類似性が比較的高く、従って、目標音声フレームの前の履歴音声フレームの対応する声門パラメータと組み合わせて目標音声フレームの対応する声門パラメータを予測することができる。一実施例において、上記した前記第１複素スペクトルに基づいて前記目標音声フレームに対して声門パラメータ予測を行い、前記目標音声フレームの対応する声門パラメータを得るステップは、さらに、前記第１複素スペクトルと前記目標音声フレームの前の履歴音声フレームの対応する声門パラメータとを第２ニューラルネットワークに入力するステップであって、前記第２ニューラルネットワークはサンプル音声フレームの対応する複素スペクトル、サンプル音声フレームの前の履歴音声フレームの対応する声門パラメータ及びサンプル音声フレームの対応する声門パラメータに基づいてトレーニングを行って得られるものである、ステップと、前記第１ニューラルネットワークによって、前記第１複素スペクトルと前記目標音声フレームの前の履歴音声

フレームの対応する声門パラメータとに基づいて前記目標音声フレームの対応する声門パラメータを出力するステップとを含む。

【００７６】

履歴音声フレームと目標音声フレームとの間に相関性があり、目標音声フレームの履歴音声フレームに対応する声門パラメータと目標音声フレームの対応する声門パラメータとの間に類似性があるため、目標音声フレームの履歴音声フレームの対応する声門パラメータを参照として、目標音声フレームの声門パラメータの予測過程に対して監視を行うことで、声門パラメータ予測の正確率を向上させることができる。

【００７７】

本願のいくつかの実施例では、音声フレームが時間的により近いほど声門パラメータの類似性がより高いため、目標音声フレームに比較的近い履歴音声フレームの対応する声門パラメータを参照とすることで、予測正確率をさらに保証することができ、たとえば、目標音声フレームの直前音声フレームの対応する声門パラメータを参照とすることができる。具体的な実施例において、参照としての履歴音声フレームの数量は１フレームであってもよく、又はマルチフレームであってもよく、具体的には、実際のニーズに応じて選択して用いることができる。

【００７８】

目標音声フレームの履歴音声フレームに対応する声門パラメータは該履歴音声フレームに対して声門パラメータ予測を行うことにより得られた声門パラメータであってもよい。換言すれば、声門パラメータの予測過程において、履歴音声フレームについて予測された声門パラメータを現在の音声フレームの声門パラメータ予測過程の参照として多重化する。

【００７９】

本実施例における第２ニューラルネットワークのトレーニング過程は、前の一実施例における第２ニューラルネットワークのトレーニング過程に類似しており、ここではトレーニングの過程を繰り返し説明しない。

【００８０】

図８は、別の一実施例に基づいて示される第２ニューラルネットワークの入力と出力の模式図である。ここで、図８における第２ニューラルネットワークの構造は図７におけるものと同じであり、図７と比べて、図８における第２ニューラルネットワークの入力は、さらに該第ｎフレームの音声フレームの直前音声フレーム（すなわち第ｎ－１フレーム）の線スペクトル周波数パラメータＬＳＦ（ｎ－１）を含む。図８に示すように、第２層のＦＣ層中に第ｎフレームの音声フレームの直前音声フレームの線スペクトル周波数パラメータＬＳＦ（ｎ－１）を埋め込んで参照情報とする。隣接する２つの音声フレームのＬＳＦパラメータの類似性が非常に高く、従って、第ｎフレームの音声フレームの履歴音声フレームの対応するＬＳＦパラメータを参照情報とすれば、ＬＳＦパラメータの予測正確率を高めることができる。

【００８１】

本願のいくつかの実施例では、上記した前記目標音声フレームの前の履歴音声フレームの対応するゲインに基づいて前記目標音声フレームに対してゲイン予測を行い、前記目標音声フレームの対応するゲインを得るステップは、さらに、前記目標音声フレームの前の履歴音声フレームの対応するゲインを第３ニューラルネットワークに入力するステップであって、前記第３ニューラルネットワークはサンプル音声フレームの前の履歴音声フレームの対応するゲインと前記サンプル音声フレームの対応するゲインとに基づいてトレーニングを行って得られるものである、ステップと、前記第３ニューラルネットワークによって、前記目標音声フレームの前の履歴音声フレームの対応するゲインに基づいて前記目標音声フレームの対応するゲインを出力するステップとを含むことができる。

【００８２】

目標音声フレームの履歴音声フレームの対応するゲインは、該第３ニューラルネットワークが該履歴音声フレームのゲイン予測を行うことにより得られるものであってもよく、換言すれば、履歴音声フレームについて予測されたゲインを目標音声フレームに対してゲ

イン予測を行う過程における第３ニューラルネットワークモデルの入力として多重化する。
【００８３】
サンプル音声フレームはサンプル音声信号に対してフレーム分割を行うことにより得られてもよく、サンプル音声信号は知られている元の音声信号と知られているノイズ信号とを組み合わせることにより得ることができる。このようにして、サンプル音声中の元の音声信号が知られている場合に、該元の音声信号に対して線形予測分析を行って、該元の音声信号を再構成することに用いられる声門パラメータ、すなわちサンプル音声フレームの対応する声門パラメータを得ることができる。
【００８４】
図９は、１つの具体的な実施例に基づいて示される第３ニューラルネットワークの模式図である。図９に示すように、第３ニューラルネットワークは１層のＬＳＴＭ層と１層のＦＣ層とを含み、ここで、ＬＳＴＭ層は１つの隠れ層であり、それは１２８個のユニットを含み、ＦＣ層の入力の次元が５１２であり、出力が１次元のゲインである。１つの具体的な実施例において、第ｎフレームの音声フレームの履歴音声フレームの対応するゲインＧ＿ｐｒｅ（ｎ）は第ｎフレームの音声フレームの最初の４つ音声フレームに対応するゲインとして定義することができ、すなわち、
Ｇ＿ｐｒｅ（ｎ）＝｛Ｇ（ｎ－１）、Ｇ（ｎ－２）、Ｇ（ｎ－３）、Ｇ（ｎ－４）｝である。
【００８５】
もちろん、ゲイン予測に用いられるものとして選択された履歴音声フレームの数量は上記のような例に限定されず、具体的には、実際のニーズに応じて選択して用いることができる。
【００８６】
上記のように示される第２ニューラルネットワークと第３ニューラルネットワークは全体的にＭ－ｔｏ－Ｎのマッピング関係（Ｎ＜＜Ｍ）を呈し、すなわち、ニューラルネットワークモデルの入力情報の次元がＭであり、出力情報の次元がＮであり、ニューラルネットワークモデルの構造を極めて大きく簡略化して、ニューラルネットワークモデルの複雑さを低減させている。
【００８７】
本願のいくつかの実施例では、上記した前記第１複素スペクトルに基づいて前記目標音声フレームに対して励起信号予測を行い、前記目標音声フレームの対応する励起信号を得るステップは、さらに、前記第１複素スペクトルを第４ニューラルネットワークに入力するステップであって、前記第４ニューラルネットワークはサンプル音声フレームの対応する複素スペクトルと前記サンプル音声フレームに対応する励起信号の周波数領域表現とに基づいてトレーニングを行って得られるものである、ステップと、前記第４ニューラルネットワークによって、前記第１複素スペクトルに基づいて前記目標音声フレームに対応する励起信号の周波数領域表現を出力するステップとを含むことができる。
【００８８】
サンプル音声フレームの対応する励起信号は、サンプル音声フレームにおける知られている元の音声信号に対して線形予測分析を行うことにより得られるものであってもよい。周波数領域表現は振幅スペクトルであってもよく、又は複素スペクトルであってもよく、ここで具体的な限定を行わない。
【００８９】
第４ニューラルネットワークをトレーニングする過程において、サンプル音声フレームの複素スペクトルを第４ニューラルネットワークモデル中に入力し、次に第４ニューラルネットワークによって、入力されたサンプル音声フレームの複素スペクトルに基づいて励起信号予測を行い、予測励起信号の周波数領域表現を出力し、次に予測励起信号の周波数領域表現と該サンプル音声フレームに対応する励起信号の周波数領域表現とに基づいて第４ニューラルネットワークのパラメータを調整する。すなわち、予測励起信号の周波数領域表現と該サンプル音声フレームに対応する励起信号の周波数領域表現との類似度が所定

の要件を満たさなければ、第４ニューラルネットワークのパラメータを調整し、第４ニューラルネットワークがサンプル音声フレームについて出力された予測励起信号の周波数領域表現と該サンプル音声フレームに対応する励起信号の周波数領域表現との間の類似度が所定の要件を満たすまで続ける。上記のようなトレーニング過程により、第４ニューラルネットワークに、音声フレームの振幅スペクトルに基づいて該音声フレームの対応する励起信号の周波数領域表現を予測する能力を学習させることができ、それにより励起信号の予測を正確に行う。

【００９０】

図１０は、１つの具体的な実施例に基づいて示される第４ニューラルネットワークの模式図である。図１０に示すように、該第４ニューラルネットワークは、１層のＬＳＴＭ層と３層のＦＣ層を含み、ここで、ＬＳＴＭ層は１つの隠れ層であり、２５６個のユニットを含み、ＬＳＴＭの入力は第ｎフレームの音声フレームの対応する第１複素スペクトルＳ’（ｎ）であり、その次元が３２１次元であってもよい。３層のＦＣ層中に含まれるユニットの数量はそれぞれ５１２、５１２及び３２１であり、最後の１層のＦＣ層は３２１次元の第ｎフレームの音声フレームに対応する励起信号の周波数領域表現Ｒ（ｎ）を出力する。入力から出力への方向に沿って、３層のＦＣ層のうちの最初の２層のＦＣ層中に活性化関数が設定され、モデルの非線形発現能力を高めることに用いられ、最後の１層のＦＣ層中に活性化関数がなく、分類出力を行うことに用いられる。

【００９１】

上記に示される第１ニューラルネットワーク、第２ニューラルネットワーク、第３ニューラルネットワーク及び第４ニューラルネットワークの構造は単に例示的なものであり、その他の実施例において、深層学習のオープンソースプラットフォーム中に相応な構造のニューラルネットワークモデルを設置し、且つ対応してトレーニングを行うこともできる。

【００９２】

本願のいくつかの実施例では、図１１に示すように、ステップ４３０は、ステップ１１１０とステップ１１２０を含み、

【００９３】

ステップ１１１０：声門フィルターにより前記目標音声フレームの対応する励起信号に対してフィルタリングを行い、フィルタリング出力信号を得る。前記声門フィルターは前記目標音声フレームの対応する声門パラメータに基づいて構築されるものである。

【００９４】

ステップ１１２０：前記目標音声フレームの対応するゲインに応じて前記フィルタリング出力信号に対して増幅処理を行い、前記目標音声フレームの対応する強調音声信号を得る。

【００９５】

声門パラメータがＬＰＣ係数であれば、直接的に上式（２）にしたがって声門フィルターの構築を行うことができる。声門フィルターがｐ次フィルターであれば、目標音声フレームの対応する声門パラメータはｐ次ＬＰＣ係数、すなわち上式（２）におけるa_1、a_2、…、a_pを含み、その他の実施例において、上式（２）における定数１はＬＰＣ係数としてもよい。

【００９６】

声門パラメータがＬＳＦパラメータであれば、ＬＳＦパラメータをＬＰＣ係数に変換し、次に対応して上式（２）にしたがって声門フィルターを構築することができる。

【００９７】

フィルタリング処理は、すなわち時間領域上の畳み込みであり、従って、上記のように声門フィルターにより励起信号に対してフィルタリングを行う過程は時間領域に変換して行うことができる。目標音声フレームに対応する励起信号の周波数領域表現を予測して得ることに加えて、励起信号の周波数領域表現を時間領域に変換し、目標音声フレームに対応する励起信号の時間領域信号を得る。

【００９８】

本願の解決手段において、目標音声フレーム中には複数のサンプル点を含む。声門フィルターにより励起信号に対してフィルタリングを行い、すなわち１つのサンプル点の前の履歴サンプル点と該声門フィルターにより畳み込みを行い、該サンプル点の対応する目標信号値を得る。

【００９９】

本願のいくつかの実施例では、前記目標音声フレームは複数のサンプル点を含み、前記声門フィルターはｐ次フィルターであり、ｐが正の整数であり、前記励起信号は前記目標音声フレームにおける複数のサンプル点のそれぞれの対応する励起信号値を含む。上記のようなフィルタリング過程に従って、ステップ１１２０は、さらに、前記目標音声フレームにおける各サンプル点の前のｐ個のサンプル点に対応する励起信号値と前記ｐ次フィルターを畳み込み、前記目標音声フレームにおける各サンプル点の目標信号値を得るステップと、時間順序に応じて前記目標音声フレームにおける全部サンプル点の対応する目標信号値を組み合わせ、前記第１音声信号を得るステップとを含む。ここで、ｐ次フィルターの表現式は上式（１）を参照することができる。つまり、目標音声フレームにおける各サンプル点に対しては、その前のｐ個のサンプル点に対応する励起信号値を利用してｐ次フィルターと畳み込みを行い、各サンプル点の対応する目標信号値を得る。

【０１００】

理解できることとして、目標音声フレームにおける最初のサンプル点に対しては、該目標音声フレームの直前音声フレームにおける最後のｐ個のサンプル点の励起信号値を借りて該最初のサンプル点の対応する目標信号値を計算する必要があり、同様に、該目標音声フレームにおける２番目のサンプル点は、目標音声フレームの直前音声フレームにおける最後の（ｐ－１）個のサンプル点の励起信号値及び目標音声フレームにおける最初のサンプル点の励起信号値とｐ次フィルターを借りて畳み込みを行って、目標音声フレームにおける２番目のサンプル点に対応する目標信号値を得る必要がある。

【０１０１】

要約すると、ステップ１１２０はさらに目標音声フレームの履歴音声フレームに対応する励起信号値の参加を必要とする。所要の履歴音声フレームにおけるサンプル点の数量は声門フィルターの次数に関連し、すなわち、声門フィルターがｐ次であれば、目標音声フレームの直前音声フレームにおける最後のｐ個のサンプル点に対応する励起信号値の参加を必要とする。

【０１０２】

関連する技術において、スペクトル推定とスペクトル回帰予測の方式で音声強調を行うことが存在する。スペクトル推定の音声強調方式は一段の混合音声に音声部分とノイズ部分が含まれると考えるため、統計モデル等によりノイズを推定することができるものであり、混合音声の対応するスペクトルからノイズの対応するスペクトルを減算すれば、残るのは音声スペクトルであり、これにより、混合音声の対応するスペクトルに基づいてノイズの対応するスペクトルを減算して得られたスペクトルはクリーンな音声信号を復元することになる。スペクトル回帰予測の音声強調方式は、ニューラルネットワークにより音声フレームの対応するマスキング閾値を予測し、該マスキング閾値は該音声フレームにおける各々の周波数点における音声成分とノイズ成分の割合を反映し、次に該マスキング閾値に基づいて混合信号スペクトルに対してゲイン制御を行い、強調された後のスペクトルを取得するということである。

【０１０３】

上記のスペクトル推定とスペクトル回帰予測による音声強調方式は、ノイズスペクトル事後確率に基づく推定であり、推定されるノイズが不正確である。たとえば、キーボード叩き等の過渡ノイズが存在する可能性があり、瞬時に発生するため、推定されるノイズスペクトルは非常に不正確であり、ノイズ抑制の効果が良くないことを引き起こす。ノイズスペクトル予測が不正確である場合に、推定されるノイズスペクトルに応じて元の混合音声信号に対して処理を行えば、混合音声信号における音声の歪みを引き起こす、又はノイズ抑制効果の劣化を引き起こす可能性があり、従って、このような状況においては、音声

忠実度とノイズ抑制との間で妥協を行う必要がある。
【０１０４】
声門パラメータ、励起信号及びゲイン予測に基づき音声強調を実現する上記実施例において、声門パラメータが音声生成の物理的過程における声門特徴と強い相関を有するため、予測された声門パラメータが目標音声フレームにおける元の音声信号の音声構造を効果的に保証し、従って、音声分解で得られた声門パラメータ、励起信号及びゲインに対して合成を行うことにより目標音声フレームの強調音声信号を得ることは、元の音声が削減されることを効果的に回避することができ、音声構造を効果的に保護し、且つ、目標音声フレームの対応する声門パラメータ、励起信号及びゲインを得た後、元のノイズ付きの音声に対して処理を行うことがなくなるため、音声忠実度とノイズ抑制との両方の間に妥協を行う必要がなくなる。
【０１０５】
図１２は、別の１つの具体的な実施例に基づいて示される音声強調方法のフローチャートである。図１２に示される実施例においては、上記第２ニューラルネットワーク、第３ニューラルネットワーク及び第４ニューラルネットワークを結合して音声分解を行う。第ｎフレームの音声フレームを目標音声フレームとすると仮定すると、該第ｎフレームの音声フレームの時間領域信号はｓ（ｎ）である。図１２に示すように、該音声強調方法はステップ１２１０～１２７０を含む。
【０１０６】
ステップ１２１０：時間周波数変換であって、第ｎフレームの音声フレームの時間領域信号ｓ（ｎ）を第ｎフレームの音声フレームの対応する複素スペクトルＳ（ｎ）に変換する。
【０１０７】
ステップ１２２０：プリエンファシスであって、複素スペクトルＳ（ｎ）に基づいて第ｎフレームの音声フレームに対してプリエンファシスを行い、第１複素スペクトルＳ’（ｎ）を得る。
【０１０８】
ステップ１２３０：第２ニューラルネットワークにより声門パラメータを予測する。該ステップにおいて、第２ニューラルネットワークの入力は第１複素スペクトルＳ’（ｎ）のみを有してもよく、第１複素スペクトルＳ’（ｎ）と該第ｎフレームの音声フレームの履歴音声フレームの対応する声門パラメータＰ＿ｐｒｅ（ｎ）とを含んでもよく、該第２ニューラルネットワークは該第ｎフレームの音声フレームの対応する声門パラメータａｒ（ｎ）を出力し、該声門パラメータはＬＰＣ係数であってもよく、ＬＳＦパラメータであってもよい。
【０１０９】
ステップ１２４０：第３ニューラルネットワークにより励起信号を予測する。第３ニューラルネットワークの入力は第１複素スペクトルＳ’（ｎ）であり、出力は該第ｎフレームの音声フレームに対応する励起信号の周波数領域表現Ｒ（ｎ）である。次にステップ１２５０によってＲ（ｎ）に対して周波数時間変換を行い、第ｎフレームの音声フレームに対応する励起信号の時間領域信号ｒ（ｎ）を得ることができる。
【０１１０】
ステップ１２６０：第４ニューラルネットワークによりゲインを予測する。第４ニューラルネットワークの入力は第ｎフレームの音声フレームの履歴音声フレームに対応するゲインＧ＿ｐｒｅ（ｎ）であり、出力は第ｎフレームの音声フレームの対応するゲインＧ（ｎ）である。
【０１１１】
第ｎフレームの音声フレームの対応する声門パラメータａｒ（ｎ）、対応する励起信号ｒ（ｎ）及び対応するゲインＧ＿（ｎ）を取得した後に、該３種のパラメータに基づきステップ１２７０で合成フィルタリングを行い、該第ｎフレームの音声フレームに対応する強調音声信号の時間領域信号ｓ＿ｅ（ｎ）を得る。ステップ１２７０の合成フィルタリン

グの過程は、図１１に示される過程を参照して行うことができる。
【０１１２】
本願の別のいくつかの実施例において、図１３に示すように、ステップ４２０は、ステップ１３１０～ステップ１３５０を含む。
【０１１３】
ステップ１３１０：前記第１複素スペクトルに基づいてパワースペクトルを計算して取得する。
【０１１４】
第１複素スペクトルがＳ’（ｎ）であれば、ステップ１３１０において得られたパワースペクトルＰａ（ｎ）は、
Ｐａ（ｎ）＝Ｒｅａｌ（Ｓ’（ｎ））２＋Ｉｍａｇ（Ｓ’（ｎ））２　（式１０）である。
【０１１５】
ここで、Ｒｅａｌ（Ｓ’（ｎ））は第１複素スペクトルＳ’（ｎ）の実部を表し、Ｉｍａｇ（Ｓ’（ｎ））は第１複素スペクトルＳ’（ｎ）の虚部を表す。ステップ１３１０において計算されて取得されたパワースペクトルは、すなわち目標音声フレームに対してプリエンファシスを行った後の信号のパワースペクトルである。
【０１１６】
ステップ１３２０：前記パワースペクトルに基づいて自己相関係数を計算して取得する。
【０１１７】
ウィナーヒンチンの定理に従う：定常なランダム過程のパワースペクトルとその自己相関関数とは一対のフーリエ変換関係である。本解決方法において、１フレームの音声フレームは定常なランダム信号と見なされる。従って、目標音声フレームに対応するプリエンファシスされた後のパワースペクトルを得たことに加えて、目標音声フレームに対応するプリエンファシスされた後のパワースペクトルに対して逆フーリエ変換を行い、該プリエンファシスされた後のパワースペクトルの対応する自己相関係数を得ることができる。
【０１１８】
具体的には、ステップ１３２０は、前記パワースペクトルに対して逆フーリエ変換を行い、逆変換結果を得て、前記逆変換結果中の実部を抽出し、前記自己相関係数を得ることを含む。すなわち、
ＡＣ（ｎ）＝Ｒｅａｌ（ｉＦＦＴ（Ｐａ（ｎ））　（式１１）
ＡＣ（ｎ）は第ｎフレームの音声フレームの対応する自己相関係数を表し、ｉＦＦＴ（Ｉｎｖｅｒｓｅ　Ｆａｓｔ　Ｆｏｕｒｉｅｒ　Ｔｒａｎｓｆｏｒｍ、逆高速フーリエ変換）とはＦＦＴ（Ｆａｓｔ　Ｆｏｕｒｉｅｒ　Ｔｒａｎｓｆｏｒｍ、高速フーリエ変換）の逆変換を指し、Ｒｅａｌは逆高速フーリエ変換で得られた結果の実部を表す。ＡＣ（ｎ）はｐ個のパラメータを含み、ｐが声門フィルターの次数であり、ＡＣ（ｎ）中の係数はさらにＡＣｊ（ｎ）として表されてもよく、１≦ｊ≦ｐである。
【０１１９】
ステップ１３３０：前記自己相関係数に基づいて前記声門パラメータを計算して取得する。
【０１２０】
Ｙｕｌｅ－Ｗａｌｋｅｒ（ユール－ウォーカー）方程式にしたがって、第ｎフレームの音声フレームに対して、その対応する自己相関係数と対応する声門パラメータとの間に以下の関係が存在する
ｋ－ＫＡ＝０　（式１２）
ここで、ｋは自己相関ベクトルであり、Ｋは自己相関行列であり、ＡはＬＰＣ係数行列である。具体的には、［数３］である。
【０１２１】
【数３】

$$r=\begin{bmatrix}AC_0(n)\\AC_1(n)\\\vdots\\AC_p(n)\end{bmatrix},\quad R=\begin{bmatrix}AC_0(n) & AC_1(n) & \cdots & AC_{p-1}(n)\\AC_1(n) & AC_0(n) & \cdots & AC_{p-2}(n)\\\vdots & \vdots & \vdots & \vdots\\AC_{p-1}(n) & AC_{p-1}(n) & \ldots & AC_0(n)\end{bmatrix}$$

$$A=\begin{bmatrix}a_1(n)\\a_2(n)\\\vdots\\a_p(n)\end{bmatrix}$$

【０１２２】

ここで、ＡＣｊ（ｎ）＝Ｅ［ｓ（ｎ）ｓ（ｎ－ｊ）］，０≦ｊ≦ｐ　（式１３）

【０１２３】

ｐは声門フィルターの次数であり、ａ１（ｎ）、ａ２（ｎ）、…、ａｐ（ｎ）はいずれも第ｎフレームの音声フレームに対応するＬＰＣ係数であり、それぞれ上式２におけるａ１、ａ２、…、ａｐであり、ａ０（ｎ）が定数１であるため、ａ０（ｎ）を第ｎフレームの音声フレームに対応する１つのＬＰＣ係数として見なすこともできる。

【０１２４】

自己相関係数を得たことに加えて、自己相関ベクトルと自己相関行列は対応して決定することができ、次に式１２を求めることにより、ＬＰＣ係数を得ることができる。具体的な実施例において、Ｌｅｖｉｎｓｏｎ－Ｄｕｒｂｉｎアルゴリズムを採用して式１２を求めることができ、Ｌｅｖｉｎｓｏｎ－Ｄｕｒｂｉｎアルゴリズムは自己相関行列の対称性を利用し、反復の方式を利用して、自己相関係数を計算して取得する。

【０１２５】

ＬＳＦパラメータとＬＰＣ係数との間は相互に変換することができ、従って、ＬＰＣ係数を計算して取得する時に、ＬＳＦパラメータを対応して決定することができる。換言すれば、声門パラメータがＬＰＣ係数であるかＬＳＦパラメータであるかにかかわらず、いずれも上記のような過程によって決定することができる。

【０１２６】

ステップ１３４０：前記声門パラメータと前記自己相関パラメータ集合とに基づいて前記ゲインを計算して取得する。

【０１２７】

以下の式［数４］にしたがって第ｎフレームの音声フレームの対応するゲインを計算することができる。

［数４］　（式１４）

【０１２８】

【数４】

$$G(n)=\sum_{j=0}^{p}AC_j(n)*a_j(n)$$

【０１２９】

式１４にしたがって計算して取得したＧ（ｎ）は時間領域表示上の目標音声フレームに対応するゲインの二乗である。

【０１３０】

ステップ１３５０：前記ゲインと声門フィルターのパワースペクトルとに基づいて前記励起信号のパワースペクトルを計算して取得する。前記声門フィルターは前記声門パラメータに基づいて構築されるフィルターである。

【０１３１】

目標音声フレームの対応する複素スペクトルがｍ（ｍが正の整数）個のサンプル点に対してはフーリエ変換を行って得られるものと仮定すると、声門フィルターのパワースペクトルを計算するためには、まず第ｎフレームの音声フレームのために次元がｍの全０の配列ｓ＿ＡＲ（ｎ）を構造し、次に、（ｐ＋１）次元のａｊ（ｎ）を該全０の配列の最初の（ｐ＋１）次元に代入し、ここでｊ＝０、１、２、…ｐであり、ｍ個のサンプル点の高速フーリエ変換（Ｆａｓｔ　Ｆｏｕｒｉｅｒ　Ｔｒａｎｓｆｏｒｍ、ＦＦＴ）を呼び出すことにより、ＦＦＴ係数を取得する。

Ｓ＿ＡＲ（ｎ）＝ＦＦＴ（ｓ＿ＡＲ（ｎ））　（式１５）

ＦＦＴ係数Ｓ＿ＡＲ（ｎ）を得たことに加えて、下式１６にしたがって１つずつのサンプルについて第ｎフレームの音声フレームに対応する声門フィルターのパワースペクトルを取得することができ、

$$AR\_LPS(n,k)=(Real(S\_AR(n,k)))^2+(Imag(S\_AR(n,k)))^2 \quad (式16)$$

ここで、Ｒｅａｌ（Ｓ＿ＡＲ（ｎ，ｋ））はＳ＿ＡＲ（ｎ，ｋ）の実部を表し、Ｉｍａｇ（Ｓ＿ＡＲ（ｎ，ｋ））はＳ＿ＡＲ（ｎ，ｋ）の虚部を表し、ｋはＦＦＴ係数の数列を表し、０≦ｋ≦ｍ、ｋは正の整数である。

【０１３２】

第ｎフレームの音声フレームに対応する声門フィルターの周波数応答ＡＲ＿ＬＰＳ（ｎ）を得た後に、計算を便利にするために、式１７にしたがって声門フィルターのパワースペクトルＡＲ＿ＬＰＳ（ｎ）を自然数領域から対数領域に変換し、

$$AR\_LPS_1(n)=\log_{10}(AR\_LPS(n)) \quad (式17)$$

上記$AR_LPS_1(n)$を下式１８にしたがって反転し、すなわち、声門フィルターの逆対応するパワースペクトル$AR_LPS_2(n)$を得て、

$$AR\_LPS_2(n)=-1*AR\_LPS_1(n) \quad (式18)$$

次に下式１９にしたがって目標音声フレームに対応する励起信号のパワースペクトルＲ（ｎ）を計算して取得することができる。

$$R(n)=Pa(n)*(G1(n))^2*AR\_LPS_3(n) \quad (式19)$$

ここで、［数５］　（式２０）

［数６］　（式２１）

【０１３３】

【数５】

$$G_1(n)=\frac{1}{\sqrt{G(n)}},$$

【数６】

$$AR\_LPS_3(n)=10^{\frac{AR\_LPS_2(n)}{10}}$$

【０１３４】

上記のような過程により、目標音声フレームに対応する声門パラメータ、ゲイン及び励起信号の周波数応答、及び声門パラメータにより限定される声門フィルターの周波数応答を計算して取得する。

【０１３５】

目標音声フレームに対応するゲイン、対応する励起信号のパワースペクトル、及び声門パラメータに限定される声門フィルターのパワースペクトルを得た後に、図１４に示される過程に基づいて合成処理を行うことができる。図１４に示すように、ステップ４３０は、ステップ１４１０～ステップ１４３０を含む。

【０１３６】

ステップ１４１０：前記声門フィルターのパワースペクトルと前記励起信号のパワースペクトルとに基づいて第１振幅スペクトルを生成する。

【０１３７】

以下の式２２にしたがって第１振幅スペクトルＳ＿ｆｉｌｔ（ｎ）を計算して取得することができる。

［数７］　（式２２）

【０１３８】

【数７】

$$S\_filt(n) = \sqrt{10^{R_1(n) + AR_1\_LPS(n)}},$$

【０１３９】

ここで、Ｒ１（ｎ）＝１０＊ｌｏｇ１０（Ｒ（ｎ））　（式２３）

【０１４０】

ステップ１４２０：前記ゲインに応じて前記第１振幅スペクトルに対して増幅処理を行い、第２振幅スペクトルを得る。

【０１４１】

下式にしたがって第２振幅スペクトルＳ＿ｅ（ｎ）を得ることができる。

Ｓ＿ｅ（ｎ）＝Ｇ２（ｎ）＊Ｓ＿ｆｉｌｔ（ｎ）　（式２４）

ここで、［数８］　（式２５）

【０１４２】

【数８】

$$G_2(n) = \sqrt{G(n)}$$

【０１４３】

ステップ１４３０：前記第２振幅スペクトルと前記第１複素スペクトル中から抽出された位相スペクトルとに基づいて、前記目標音声フレームの対応する強調音声信号を決定する。

【０１４４】

本願のいくつかの実施例では、ステップ１４３０は、さらに、前記第２振幅スペクトルと前記第１複素スペクトル中から抽出された位相スペクトルとを組み合わせ、第２複素スペクトルを得るステップ、換言すれば、第２振幅スペクトルを第２複素スペクトルの実部とし、第１複素スペクトル中から抽出された位相スペクトルを第２複素スペクトルの虚部とし、前記第２複素スペクトルを時間領域に変換し、前記目標音声フレームに対応する強調音声信号の時間領域信号を得るステップを含む。

【０１４５】

図１５は、１つの具体的な実施例に基づいて示される音声強調方法のフローチャートであり、第ｎフレームの音声フレームを目標音声フレームとし、第ｎフレームの音声フレームの時間領域信号がｓ（ｎ）である。図１５に示すように、具体的には、ステップ１５１０～１５６０を含む。

【０１４６】

ステップ１５１０：時間周波数変換であって、ステップ１５１０により第ｎフレームの音声フレームの時間領域信号ｓ（ｎ）を変換して第ｎフレームの音声フレームの対応する複素スペクトルＳ（ｎ）を得る。

【０１４７】

ステップ１５２０：プリエンファシスであって、第ｎフレームの音声フレームの対応する複素スペクトルＳ（ｎ）に基づき該第ｎフレームの音声フレームに対してプリエンファシス処理を行い、第ｎフレームの音声フレームの第１複素スペクトルＳ′（ｎ）を得る。

【０１４８】

ステップ１５３０：スペクトル分解であって、第１複素スペクトルＳ′（ｎ）に対してスペクトル分解を行うことにより、第１複素スペクトルＳ′（ｎ）の対応するパワースペクトルＰａ（ｎ）と対応する位相スペクトルＰｈ（ｎ）とを得る。

【０１４９】

ステップ１５４０：音声分解であって、第ｎフレームの音声フレームのパワースペクトルＰａ（ｎ）に基づき音声分解を行い、第ｎフレームの音声フレームの対応する声門パラメータ集合Ｐ（ｎ）と第ｎフレームの音声フレームに対応する励起信号の周波数領域表現Ｒ（ｎ）とを決定する。声門パラメータ集合Ｐ（ｎ）は声門パラメータａｒ（ｎ）とゲインＧ（ｎ）を含む。具体的な音声分解の過程は図１３に示されてもよく、声門パラメータを取得し、且つ声門フィルターのパワースペクトルＡＲ＿ＬＰＳ（ｎ）、励起信号のパワースペクトルＲ（ｎ）、及びゲインＧ（ｎ）を対応して取得する。

【０１５０】

ステップ１５５０：音声合成する。具体的な音声合成の過程は図１４に示されてもよく、第ｎフレームの音声フレームに対応する声門フィルターの周波数応答ＡＲ＿ＬＰＳ（ｎ）、励起信号の周波数応答Ｒ（ｎ）、及びゲインＧ（ｎ）に対して合成を行って第２振幅スペクトルＳ＿ｅ（ｎ）を得る。

【０１５１】

ステップ１５６０：周波数時間変換する。第１複素スペクトルＳ′（ｎ）から抽出された位相スペクトルＰｈ（ｎ）を多重化し、位相スペクトルＰｈ（ｎ）と第２振幅スペクトルＳ＿ｅ（ｎ）を組み合わせて第ｎフレームの音声フレームに対応する強調された後の複素スペクトルを得る。得られた強調された後の複素スペクトルを時間領域に変換すると、第ｎフレームの音声フレームに対応する強調音声信号の時間領域信号ｓ＿ｅ（ｎ）を得る。

【０１５２】

本実施例の解決手段において、目標音声フレームに対してプリエンファシスを行うことにより得られた第１複素スペクトルに基づいて音声分解を行い、プリエンファシスする過程において、一部のノイズの情報が除外され、従って、第１複素スペクトルにおけるノイズ情報がより少なくなる。従って、第１複素スペクトルに基づいて音声分解を行うことで、ノイズによる音声分解への影響を減少し、音声分解の難度を低減させ、音声分解で得られた声門パラメータ、励起信号及びゲインの正確性を向上させ、さらに後続で取得された強調音声信号の正確性を保証することができる。また、本解決方法において、音声合成過程において、振幅スペクトルのみに注目することができ、位相情報に注目する必要がなく、第１複素スペクトル中から抽出された位相スペクトルを直接的に多重化することにより、音声合成過程における計算量を減少させる。第１複素スペクトルはプリエンファシスを行って得られるものであり、そのノイズ含有量がより少なく、従って、ある程度で位相情報の精度を保証する。

【０１５３】

図１５に示される実施例においては、ステップ１５１０において、第１ニューラルネットワークによってプリエンファシスを実現することができる。ステップ１５４０は図１３に示される過程にしたがって実現でき、ステップ１５５０は図１４に示される過程にしたがって実現でき、それにより、従来信号処理と深層学習とを深く組み合わせ、且つ目標音声フレームに対して二次強調を行うことが実現される。従って、本願の実施例は目標音声フレームに対して複数段階の強調を行うことを実現する。すなわち、第１段階では、深層学習の方式を採用して目標音声フレームの振幅スペクトルに基づいてプリエンファシスを行い、第２段階における音声分解して声門パラメータ、励起信号及びゲインを取得する難しさを低減させることができ、第２段階では、信号処理の方式により元の音声信号を再構成することに用いられる声門パラメータ、励起信号及びゲインを取得する。そして、第２段階において、音声が生じているデジタルモデルにしたがって音声合成を行い、目標音声フレームの信号に対して処理を直接的に行わず、従って、第２段階における音声削減状況の出現を回避することができる。

【０１５４】

本願のいくつかの実施例では、ステップ４１０の前に、該方法は、さらに、前記目標音声フレームの時間領域信号を取得するステップと、前記目標音声フレームの時間領域信号に対して時間周波数変換を行い、前記目標音声フレームの複素スペクトルを得るステップとを含む。

【０１５５】

時間周波数変換は短時間フーリエ変換（ｓｈｏｒｔ－ｔｅｒｍ Ｆｏｕｒｉｅｒ ｔｒａｎｓｆｏｒｍ、ＳＴＦＴ）であってもよい。短時間フーリエ変換において窓掛け・オーバーラップの操作を採用してフレームの間の不平滑化を解消する。図１６は１つの具体的な実施例に基づいて示される短時間フーリエ変換における窓掛け・オーバーラップの模式図であり、図１６において、５０％窓掛け・オーバーラップの操作を採用し、短時間フーリエ変換が６４０個のサンプル点に対するものであれば、該窓関数の重なったサンプル数（ｈｏｐ－ｓｉｚｅ）は３２０である。窓掛けに使用される窓関数はハニング（Ｈａｎｎｉｎｇ）窓、ハミング窓等であってもよく、もちろん、その他の窓関数を採用してもよく、ここで具体的な限定を行わない。

【０１５６】

その他の実施例において、５０％ではない窓掛け・オーバーラップの操作を採用してもよい。たとえば、短時間フーリエ変換が５１２個のサンプル点に対するものであれば、この場合には、１つの音声フレーム中に３２０個のサンプル点が含まれれば、直前音声フレームの１９２個のサンプル点をオーバーラップするだけでよい。

【０１５７】

本願のいくつかの実施例では、目標音声フレームの時間領域信号を取得するステップは、さらに、処理対象の音声信号を取得するステップであって、前記処理対象の音声信号は収集された音声信号又は符号化音声に対して復号を行って得られた音声信号である、ステップと、前記処理対象の音声信号に対してフレーム分割を行い、前記目標音声フレームの時間領域信号を得るステップとを含む。

【０１５８】

いくつかの実例において、設定されたフレーム長さに応じて処理対象の音声信号に対してフレーム分割を行うことができ、該フレーム長さは実際のニーズに応じて設定を行うことができ、たとえば、フレーム長さが２０ｍｓに設定される。フレーム分割を行うことにより、複数の音声フレームを得ることができ、各音声フレームはいずれも本願における目標音声フレームとすることができる。

【０１５９】

上記の記述のように、本願の解決手段は送信端に適用され音声強調を行うことができ、受信端に適用され音声強調を行うこともできる。本願の解決手段が送信端に適用される場合に、該処理対象の音声信号は送信端が収集した音声信号であり、その場合、処理対象の音声信号に対してフレーム分割を行い、複数の音声フレームを得る。フレーム分割の後、処理対象の音声信号は複数の音声フレームに分割され、次に各音声フレームを目標音声フレームとし且つ上記ステップ４１０～４３０の過程にしたがって目標音声フレームに対して強調を行うことができる。さらには、目標音声フレームの対応する強調音声信号を得た後に、さらに該強調音声信号に対して符号化を行うこともでき、それにより、得られた符号化に基づき音声伝送を行う。

【０１６０】

一実施例において、直接収集された音声信号はアナログ信号であるため、信号処理を便利に行うために、フレーム分割を行う前に、音声信号をさらにデジタル化し、時間的に連続する音声信号を時間的に離散する音声信号に変換する必要もある。デジタル化を行う過程において、設定されたサンプリングレートに応じて収集された音声信号に対してサンプリングを行うことができ、設定されたサンプリングレートは１６０００Ｈｚ、８０００Ｈｚ、３２０００Ｈｚ、４８０００Ｈｚ等であってもよく、具体的には、実際のニーズに応じて設定を行うことができる。

【０１６１】

本願の解決手段が受信端に適用される場合に、該処理対象の音声信号は受信された符号化音声に対して復号を行って得られた音声信号である。このような場合に、送信端が、伝送する必要がある音声信号に対して強調を行っていない可能性があり、従って、信号品質を向上させるためには、受信端で音声信号に対して強調を行う必要がある。処理対象の音声信号に対してフレーム分割を行って複数の音声フレームを得た後に、それを目標音声フレームとし、且つ上記のようなステップ４１０～４３０の過程にしたがって目標音声フレームに対して強調を行い、目標音声フレームの強調音声信号を得る。さらに、目標音声フレームの対応する強調音声信号に対して再生を行うこともでき、得られた強調音声信号は目標音声フレームの強調前の信号に比べて、ノイズが既に除去されているため、音声信号の品質がより高く、従って、ユーザーにとって、聴覚的体験がより高い。

【０１６２】

以下、本願の上記実施例における方法を実行することに用いることができる本願の装置の実施例を説明する。本願の装置実施例において披露されない細部に対しては、本願の上記方法実施例を参照されたい。

【０１６３】

図１７は、一実施例に基づいて示される音声強調装置のブロック図である。図１７に示すように、該音声強調装置は、目標音声フレームの複素スペクトルに基づいて前記目標音声フレームに対してプリエンファシス処理を行い、第１複素スペクトルを得ることに用いられるプリエンファシスモジュール１７１０と、前記第１複素スペクトルに基づいて前記目標音声フレームに対して音声分解を行い、前記目標音声フレームの対応する声門パラメータ、ゲイン及び励起信号を得ることに用いられる音声分解モジュール１７２０と、前記声門パラメータ、前記ゲイン及び前記励起信号に基づいて合成処理を行い、前記目標音声フレームの対応する強調音声信号を得ることに用いられる合成処理モジュール１７３０とを含む。

【０１６４】

本願のいくつかの実施例では、プリエンファシスモジュール１７１０は、前記目標音声フレームの対応する複素スペクトルを第１ニューラルネットワークに入力することに用いられる第１入力ユニットであって、前記第１ニューラルネットワークはサンプル音声フレームの対応する複素スペクトルと前記サンプル音声フレームにおける元の音声信号の対応する複素スペクトルとに基づいてトレーニングを行って得られるものである、第１入力ユニットと、前記第１ニューラルネットワークによって、前記目標音声フレームの対応する複素スペクトルに基づいて前記第１複素スペクトルを出力することに用いられる第１出力ユニットとを含む。

【０１６５】

本願のいくつかの実施例では、前記第１ニューラルネットワークは複素畳み込み層、ゲート付き回帰型ユニット層及び全結合層を含み、第１出力ユニットは、前記複素畳み込み層によって前記目標音声フレームに対応する複素スペクトルにおける実部及び虚部に基づいて複素畳み込み処理を行うことに用いられる複素畳み込みユニットと、前記ゲート付き回帰型ユニット層によって前記複素畳み込み層の出力に対して変換処理を行うことに用いられる変換ユニットと、前記全結合層によって前記ゲート付き回帰型ユニットの出力に対して全結合処理を行い、前記第１複素スペクトルを出力することに用いられる全結合ユニットとを含む。

【０１６６】

本願のいくつかの実施例では、音声分解モジュール１７２０は、前記第１複素スペクトルに基づいて前記目標音声フレームに対して声門パラメータ予測を行い、前記目標音声フレームの対応する声門パラメータを得ることに用いられる声門パラメータ予測ユニットに用いられる第１振幅スペクトル取得ユニットと、前記第１複素スペクトルに基づいて前記目標音声フレームに対して励起信号予測を行い、前記目標音声フレームの対応する励起信号を得ることに用いられる励起信号予測ユニットと、前記目標音声フレームの前の履歴音声フレームの対応するゲインに基づいて前記目標音声フレームに対してゲイン予測を行い

、前記目標音声フレームの対応するゲインを得ることに用いられるゲイン予測ユニットとを含む。

【０１６７】

本願のいくつかの実施例では、声門パラメータ予測ユニットは、前記第１複素スペクトルを第２ニューラルネットワークに入力することに用いられる第２入力ユニットであって、前記第２ニューラルネットワークはサンプル音声フレームの対応する複素スペクトルと前記サンプル音声フレームの対応する声門パラメータとに基づいてトレーニングを行って得られるものである、第２入力ユニットと、前記第２ニューラルネットワークによって、前記第１複素スペクトルに基づいて前記目標音声フレームの対応する声門パラメータを出力することに用いられる第２出力ユニットとを含む。

【０１６８】

本願の別のいくつかの実施例において、声門パラメータ予測ユニットは、前記第１複素スペクトルと前記目標音声フレームの前の履歴音声フレームの対応する声門パラメータとを第２ニューラルネットワークに入力することに用いられる第３入力ユニットであって、前記第２ニューラルネットワークはサンプル音声フレームの対応する複素スペクトル、サンプル音声フレームの前の履歴音声フレームの対応する声門パラメータ及びサンプル音声フレームの対応する声門パラメータに基づいてトレーニングを行って得られるものである、第３入力ユニットと、前記第１ニューラルネットワークによって、前記第１複素スペクトルと前記目標音声フレームの前の履歴音声フレームの対応する声門パラメータとに基づいて前記目標音声フレームの対応する声門パラメータを出力することに用いられる第３出力ユニットとを含む。

【０１６９】

本願のいくつかの実施例では、ゲイン予測ユニットは、前記目標音声フレームの前の履歴音声フレームの対応するゲインを第３ニューラルネットワークに入力することに用いられる第４入力ユニットであって、前記第３ニューラルネットワークはサンプル音声フレームの前の履歴音声フレームの対応するゲインと前記サンプル音声フレームの対応するゲインとに基づいてトレーニングを行って得られるものである、第４入力ユニットと、前記第３ニューラルネットワークによって、前記目標音声フレームの前の履歴音声フレームの対応するゲインに基づいて前記目標音声フレームの対応するゲインを出力することに用いられる第４出力ユニットとを含む。

【０１７０】

本願のいくつかの実施例では、励起信号予測ユニットは、前記第１複素スペクトルを第４ニューラルネットワークに入力することに用いられる第５入力ユニットであって、前記第４ニューラルネットワークはサンプル音声フレームの対応する複素スペクトルと前記サンプル音声フレームに対応する励起信号の周波数領域表現とに基づいてトレーニングを行って得られるものである、第５入力ユニットと、前記第４ニューラルネットワークによって、前記第１複素スペクトルに基づいて前記目標音声フレームに対応する励起信号の周波数領域表現を出力することに用いられる第５出力ユニットとを含む。

【０１７１】

本願のいくつかの実施例では、合成処理モジュール１７３０は、声門フィルターにより前記目標音声フレームの対応する励起信号に対してフィルタリングを行い、フィルタリング出力信号を得ることに用いられるフィルタリングユニットであって、前記声門フィルターは前記目標音声フレームの対応する声門パラメータに基づいて構築されるものである、フィルタリングユニットと、前記目標音声フレームの対応するゲインに応じて前記フィルタリング出力信号に対して増幅処理を行い、前記目標音声フレームの対応する強調音声信号を得ることに用いられる増幅処理ユニットとを含む。

【０１７２】

本願のいくつかの実施例では、音声分解モジュール１７２０は、前記第１複素スペクトルに基づいてパワースペクトルを計算して取得することに用いられるパワースペクトル計算ユニットと、前記パワースペクトルに基づいて自己相関係数を計算して取得することに

用いられる自己相関係数計算ユニットと、前記自己相関係数に基づいて前記声門パラメータを計算して取得することに用いられる声門パラメータ計算ユニットと、前記声門パラメータと前記自己相関パラメータ集合とに基づいて前記ゲインを計算して取得することに用いられるゲイン計算ユニットと、前記ゲインと声門フィルターのパワースペクトルとに基づいて前記励起信号のパワースペクトルを計算して取得することに用いられる励起信号決定ユニットであって、前記声門フィルターは前記声門パラメータに基づいて構築されるフィルターである、励起信号決定ユニットとを含む。

【０１７３】

本願のいくつかの実施例では、合成処理モジュール１７３０は、前記声門フィルターのパワースペクトルと前記励起信号のパワースペクトルとに基づいて第１振幅スペクトルを生成することに用いられる第２振幅スペクトル生成ユニットと、前記ゲインに応じて前記第１振幅スペクトルに対して増幅処理を行い、第２振幅スペクトルを得ることに用いられる第３振幅スペクトル決定ユニットと、前記第２振幅スペクトルと前記第１複素スペクトル中から抽出された位相スペクトルとに基づいて、前記目標音声フレームの対応する強調音声信号を決定することに用いられる強調音声信号決定ユニットとを含む。

【０１７４】

本願のいくつかの実施例では、強調音声信号決定ユニットは、前記第２振幅スペクトルと前記第１複素スペクトル中から抽出された位相スペクトルとを組み合わせ、第２複素スペクトルを得ることに用いられる第２複素スペクトル計算ユニットと、前記第２複素スペクトルを時間領域に変換し、前記目標音声フレームに対応する強調音声信号の時間領域信号を得ることに用いられる時間領域変換ユニットとを含む。

【０１７５】

図１８は、本願の実施例を実現するための電子機器に適するコンピュータシステムの構造模式図を示す。

【０１７６】

説明する必要があることとして、図１８に示される電子機器のコンピュータシステム１８００は一例に過ぎず、本願の実施例の機能及び使用範囲に対して何ら制限をもたらすべきではない。

【０１７７】

図１８に示すように、コンピュータシステム１８００は中央処理ユニット（Ｃｅｎｔｒａｌ Ｐｒｏｃｅｓｓｉｎｇ Ｕｎｉｔ、ＣＰＵ）１８０１を含み、それは読み出し専用メモリ（Ｒｅａｄ－Ｏｎｌｙ Ｍｅｍｏｒｙ、ＲＯＭ）１８０２において記憶されたプログラム又は記憶部分１８０８からランダムアクセスメモリ（Ｒａｎｄｏｍ Ａｃｃｅｓｓ Ｍｅｍｏｒｙ、ＲＡＭ）１８０３中にアップロードされたプログラムに基づいて各種の適当な動作と処理を実行することができ、たとえば、上記実施例における方法を実行する。ＲＡＭ １８０３において、システム操作に必要な各種のプログラムとデータも記憶されている。ＣＰＵ１８０１、ＲＯＭ１８０２及びＲＡＭ １８０３はバス１８０４を介して互いに連結される。入力／出力（Ｉｎｐｕｔ／Ｏｕｔｐｕｔ、Ｉ／Ｏ）インターフェース１８０５もバス１８０４に接続される。

【０１７８】

以下の部材がＩ／Ｏインターフェース１８０５に接続される。キーボード、マウス等を含む入力部分１８０６、陰極線管（Ｃａｔｈｏｄｅ Ｒａｙ Ｔｕｂｅ、ＣＲＴ）、液晶ディスプレイ（Ｌｉｑｕｉｄ Ｃｒｙｓｔａｌ Ｄｉｓｐｌａｙ、ＬＣＤ）等のようなもの及びスピーカ等を含む出力部分１８０７、ハードディスク等を含む記憶部分１８０８、及びＬＡＮ（Ｌｏｃａｌ Ａｒｅａ Ｎｅｔｗｏｒｋ、ローカルエリアネットワーク）カード、モデム等のようなネットワークインタフェースカードを含む通信部分１８０９である。通信部分１８０９は、インターネットのようなネットワークを介して通信処理を実行する。ドライバ１８１０もニーズに応じてＩ／Ｏインターフェース１８０５に接続される。着脱可能な媒体１８１１、例えば磁気ディスク、光ディスク、光磁気ディスク、半導体メモリ等は、ニーズに応じてドライバ１８１０上に装着され、それにより、その上から読み出さ

れたコンピュータプログラムがニーズに応じて記憶部分１８０８にインストールされる。
【０１７９】
特に、本願の実施例に基づき、上記のフローチャートを参照して記述される過程はコンピュータソフトウェアプログラムとして実現できる。たとえば、本願の実施例は、１種のコンピュータプログラム製品を含み、それはコンピュータ可読媒体上に担持されるコンピュータプログラムを含み、該コンピュータプログラムはフローチャートに示された方法を実行することに用いられるプログラムコードを含む。このような実施例において、該コンピュータプログラムは通信部分１８０９によりネットワーク上からダウンロードされインストールすることができ、且つ／又は着脱可能な媒体１８１１からインストールされる。該コンピュータプログラムが中央処理ユニット（ＣＰＵ）１８０１によって実行されるときに、本願のシステム中に限定される各種の機能を実行する。
【０１８０】
説明する必要があることとして、本願の実施例に示されるコンピュータ可読媒体はコンピュータ可読信号媒体、又はコンピュータ可読記憶媒体又は上記両方の任意の組み合わせであってもよい。コンピュータ可読記憶媒体は、たとえば、電気、磁気、光、電磁、赤外線、又は半導体のシステム、装置又はデバイス、又は以上の任意の組み合わせであってもよいがこれらに限定されない。コンピュータ可読記憶媒体のより具体的な例は、１つ又は複数の導線を有する電気的接続、ポータブルコンピュータ磁気ディスク、ハードディスク、ランダムアクセスメモリ（ＲＡＭ）、読み出し専用メモリ（ＲＯＭ）、消去可能なプログラマブル読み出し専用メモリ（Ｅｒａｓａｂｌｅ　Ｐｒｏｇｒａｍｍａｂｌｅ　Ｒｅａｄ　Ｏｎｌｙ　Ｍｅｍｏｒｙ、ＥＰＲＯＭ）、フラッシュメモリ、光ファイバー、ポータブルコンパクト磁気ディスク読み出し専用メモリ（Ｃｏｍｐａｃｔ　Ｄｉｓｃ　Ｒｅａｄ－Ｏｎｌｙ　Ｍｅｍｏｒｙ、ＣＤ－ＲＯＭ）、光記憶デバイス、磁気記憶デバイス、又は上記の任意の適切な組み合わせを含んでもよいがこれらに限定されない。本願において、コンピュータ可読記憶媒体は、プログラムを含む又は記憶する任意の有形媒体であってもよく、該プログラムは指令実行システム、装置又はデバイスに使用され又はそれと組み合わせて使用することができる。本願において、コンピュータ可読の信号媒体は、ベースバンド中における又は搬送波の一部として伝播されるデータ信号を含んでもよく、その中でコンピュータ可読のプログラムコードが担持されている。このような伝播されるデータ信号は複数種の形式を採用することができ、電磁信号、光信号又は上記任意の適切な組み合わせを含むがこれらに限定されない。コンピュータ可読の信号媒体はさらにコンピュータ可読記憶媒体以外の任意のコンピュータ可読媒体であってもよく、該コンピュータ可読媒体は、指令実行システム、装置又はデバイスに使用され又はそれと組み合わせて使用されることに用いられるプログラムを送信、伝播又は伝送することができる。コンピュータ可読媒体上に含まれるプログラムコードは任意の適当な媒体で伝送でき、無線、有線等、又は上記の任意の適切な組み合わせを含むがこれらに限定されない。
【０１８１】
図面におけるフローチャートとブロック図は、本願の各種の実施例に係るシステム、方法及びコンピュータプログラム製品の実現可能な体系アーキテクチャ、機能及び操作を図示する。ここで、フローチャート又はブロック図における各ブロックは１つのモジュール、プログラムセグメント、又はコードの一部を代表することができ、上記モジュール、プログラムセグメント、又はコードの一部は規定されるロジック機能を実現することに用いられる１つ又は複数の実行可能な指示を含む。また、注意すべきことは、代替としてのいくつかの実現形式において、ブロック中にマークされる機能は図面中にマークされる順序と異なるものとして生じさせることができる点である。たとえば、連続的に示される２つのブロックは実際には基本的に並行して実行することができ、場合によって、それらは逆の順序で実行することもでき、これは関連する機能によって定められる。また注意する必要があるのは、ブロック図又はフローチャートにおける各ブロック、及びブロック図又はフローチャートにおけるブロックの組み合わせは、規定される機能又は操作を実行する専用のハードウェアに基づくシステムで実現することができ、又は専用ハードウェアとコン

ピュータ指令の組み合わせで実現することもできる。

【０１８２】

本願の実施例においてに記述されて言及されるユニットはソフトウェアの方式で実現されても、又はハードウェアの方式で実現されてもよく、記述されるユニットはプロセッサ中に設置されてもよい。ここで、これらのユニットの名称がある場合には、該ユニット自体に対する限定を構成しない。

【０１８３】

別の態様として、本願はコンピュータ可読記憶媒体をさらに提供し、該コンピュータ可読媒体は上記実施例に記述される電子機器に含まれてもよく、単独で存在し、該電子機器中に組み立てられなくてもよい。上記コンピュータ可読記憶媒体はコンピュータ可読指令を担持し、該コンピュータ可読記憶指令がプロセッサによって実行されるときに、上記いずれかの実施例における方法を実現する。

【０１８４】

本願の一態様によれば、電子機器をさらに提供し、それは、プロセッサと、メモリであって、メモリ上にコンピュータ可読指令が記憶され、コンピュータ可読指令がプロセッサによって実行されるときに、上記いずれかの実施例における方法を実現するメモリとを含む。

【０１８５】

本願の実施例の一態様によれば、コンピュータプログラム製品、又はコンピュータプログラムを提供し、該コンピュータプログラム製品、又はコンピュータプログラムはコンピュータ指令を含み、該コンピュータ指令がコンピュータ可読記憶媒体中に記憶される。コンピュータ機器のプロセッサはコンピュータ可読記憶媒体から該コンピュータ指令を読み取り、プロセッサは該コンピュータ指令を実行し、該コンピュータ機器に上記いずれかの実施例における方法を実行させる。

【０１８６】

注意すべきことは、上記詳細な記述において動作実行用の機器の複数のモジュール又はユニットが言及されているが、このような分割は強制的ではないことである。実際には、本願の実施形態によれば、上記で記述された２つ又はより多くのモジュール又はユニットの特徴と機能は１つのモジュール又はユニットにおいて具現化され得る。逆に、上記で記述された１つのモジュール又はユニットの特徴と機能はさらに複数のモジュール又はユニットにより具現化されるように分割されてもよい。

【０１８７】

以上の実施形態の記述により、当業者が容易に理解できることは、ここで記述される例示的な実施形態はソフトウェアで実現されてもよく、ソフトウェアと必要なハードウェアを組み合わせた方式で実現されてもよい。従って、本願の実施形態に係る技術的手段は、ソフトウェア製品の形式で体現されてもよく、該ソフトウェア製品は１つの不揮発性記憶媒体（ＣＤ－ＲＯＭ、Ｕディスク、モバイルディスク等であってもよい）中に又はネットワーク上に記憶されてもよく、幾つかの指令を含むことで一台の計算機器（パソコンコンピュータ、サーバ、タッチ端末、又はネットワーク機器等であってもよい）に本願の実施形態に係る方法を実行させる。

【０１８８】

当業者は明細書を考慮し、且つここで開示される実施形態を実践した後に、本願のその他の実施形態を容易に想到することができる。本願は本願の任意の変形、用途又は適応的な変化をカバーすることを目的としており、これらの変形、用途又は適応的な変化は本願の一般原理に従い、且つ本願に開示されていない本技術分野における公知の知識又は一般的な技術手段を含む。

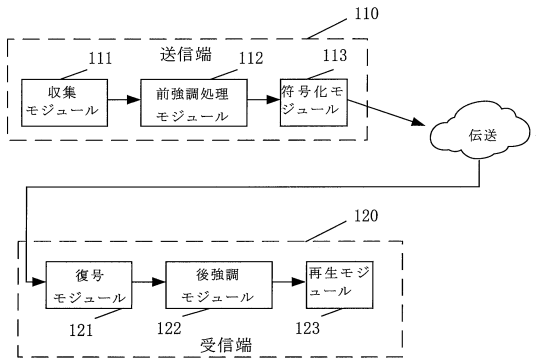
【０１８９】

理解すべきことは、本願は上記において記述され、且つ図面中に示される正確な構造には限定されず、且つその範囲を逸脱することなく、各種の修正や変更を行うことができる。本願の範囲は添付の請求項の記載のみによって制限される。

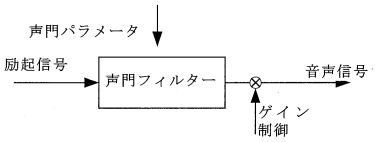
【符号の説明】
【０１９０】
１１０　送信端
１１１　収集モジュール
１１２　前強調処理モジュール
１１３　符号化モジュール
１２０　受信端
１２１　復号モジュール
１２２　後強調モジュール
１２３　再生モジュール
１７１０　プリエンファシスモジュール
１７２０　音声分解モジュール
１７３０　合成処理モジュール
１８００　コンピュータシステム
１８０１　中央処理ユニット（ＣＰＵ）
１８０４　バス
１８０５　Ｉ／Ｏインターフェース
１８０５　出力（Ｉｎｐｕｔ／Ｏｕｔｐｕｔ、Ｉ／Ｏ）インターフェース
１８０６　入力部分
１８０７　出力部分
１８０８　記憶部分
１８０９　通信部分
１８１０　ドライバ
１８１１　媒体

【図面】

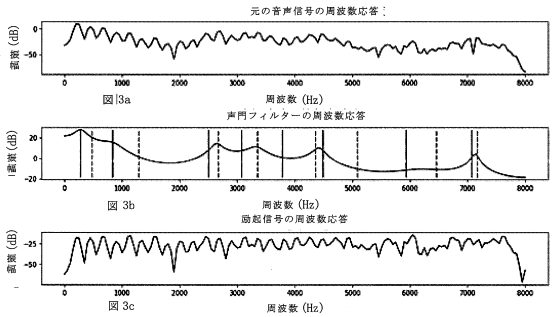
【図１】


110
送信端
111
収集
モジュール
112
前強調処理
モジュール
113
符号化モ
ジュール
伝送
120
復号
モジュール
後強調
モジュール
再生モジ
ュール
121
122
123
受信端

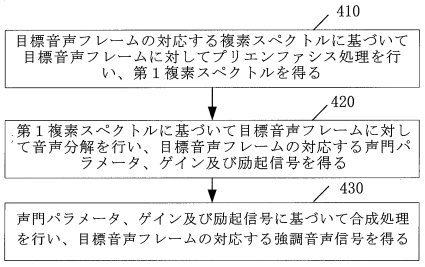
【図２】


声門パラメータ
励起信号
声門フィルター
音声信号
ゲイン
制御

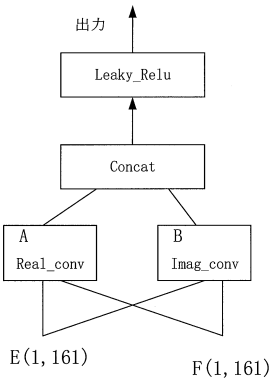
【図３】


元の音声信号の周波数応答
音量 (dB)
周波数 (Hz)
図 3a
声門フィルターの周波数応答
音量 (dB)
周波数 (Hz)
図 3b
励起信号の周波数応答
音量 (dB)
周波数 (Hz)
図 3c

【図４】


410
目標音声フレームの対応する複素スペクトルに基づいて目標音声フレームに対してプリエンファシス処理を行い、第１複素スペクトルを得る
420
第１複素スペクトルに基づいて目標音声フレームに対して音声分解を行い、目標音声フレームの対応する声門パラメータ、ゲイン及び励起信号を得る
430
声門パラメータ、ゲイン及び励起信号に基づいて合成処理を行い、目標音声フレームの対応する強調音声信号を得る

【図５】



【図６】

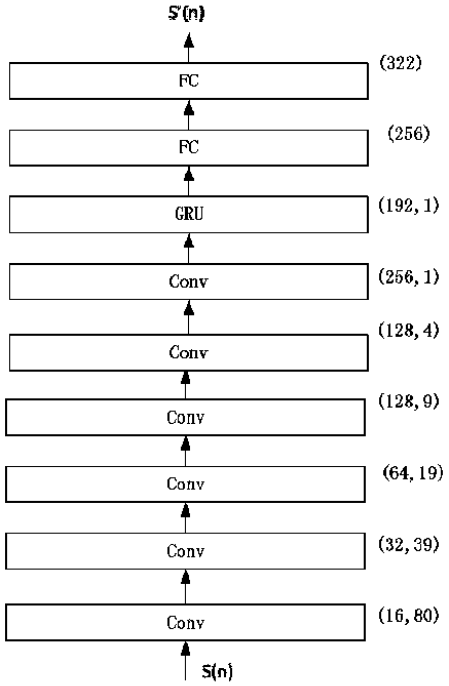


図 6

【図７】

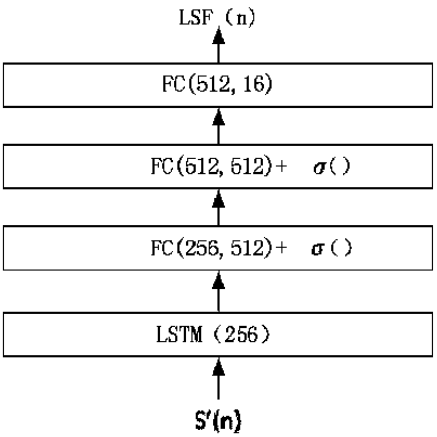


図 7

【図８】

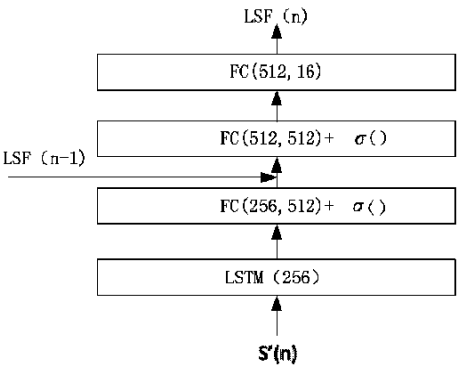


図 8

【図９】

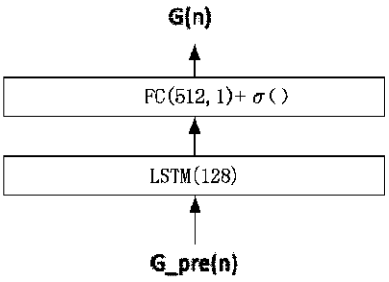


図 9

【図１０】

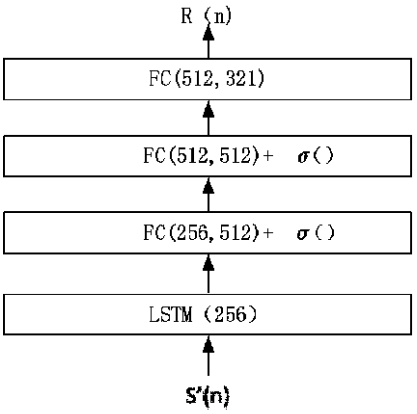
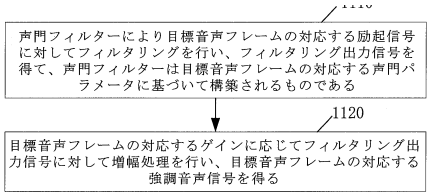
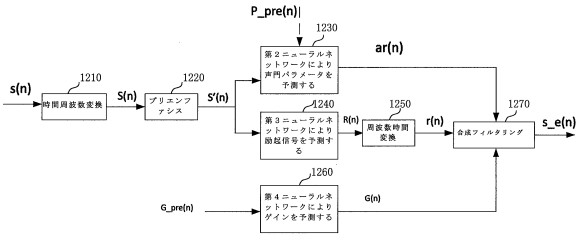


図 10

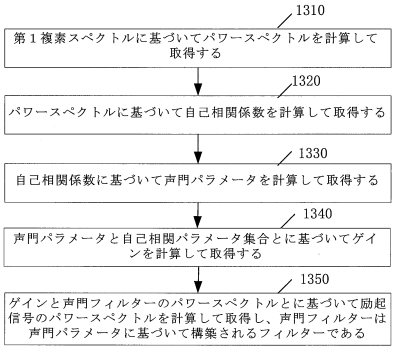
【図１１】



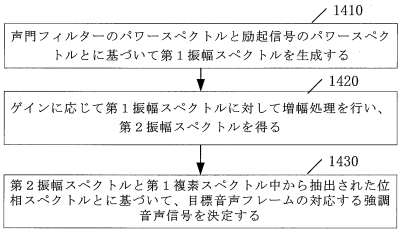
【図１２】



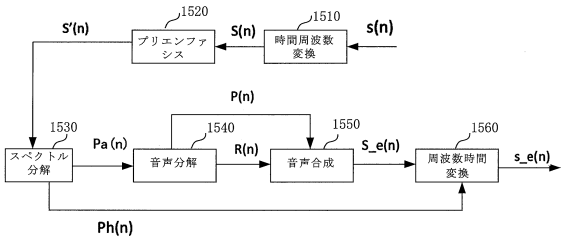
【図１３】



【図１４】



【図１５】



【図１６】

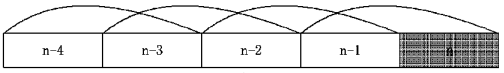
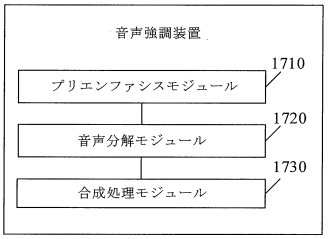
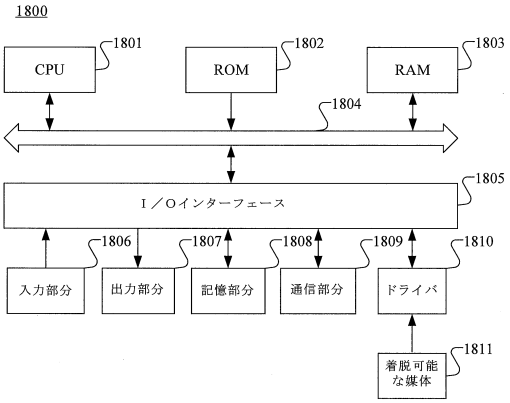


图 16

【図１７】



【図１８】



フロントページの続き

０５７，ＣＨＩＮＡ
(74)代理人　100110364
弁理士　実広 信哉
(74)代理人　100150197
弁理士　松尾 直樹
(72)発明者　肖 ▲ウェイ▼
中華人民共和国５１８０５７▲広▼▲東▼省深▲セン▼市南山区高新区科技中一路 ▲騰▼▲訊▼大厦３５▲層▼
(72)発明者　史 裕▲鵬▼
中華人民共和国５１８０５７▲広▼▲東▼省深▲セン▼市南山区高新区科技中一路 ▲騰▼▲訊▼大厦３５▲層▼
(72)発明者　王 蒙
中華人民共和国５１８０５７▲広▼▲東▼省深▲セン▼市南山区高新区科技中一路 ▲騰▼▲訊▼大厦３５▲層▼

審査官　中村 天真
(56)参考文献　特開２０００－３４７６９８（ＪＰ，Ａ）
特開２０２０－０６０６１２（ＪＰ，Ａ）
特開平０２－１３７９００（ＪＰ，Ａ）
特開２０２０－１２２８９６（ＪＰ，Ａ）
特開平１０－１９０４９８（ＪＰ，Ａ）
特開２００２－０４１０８５（ＪＰ，Ａ）
特開２００２－３６６２００（ＪＰ，Ａ）
米国特許第０５１４８４８８（ＵＳ，Ａ）
(58)調査した分野　(Int.Cl.，ＤＢ名)
Ｇ１０Ｌ １３／００－１３／１０
１９／００－９９／００