(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2014-199445
(P2014-199445A)
(43)公開日　平成26年10月23日(2014.10.23)

(51)Int.Cl.		F I		テーマコード（参考）
G10L 21/02	(2013.01)	G10L 21/02		5D061
G10K 11/178	(2006.01)	G10K 11/16	H	

審査請求 未請求　請求項の数 13　ＯＬ　（全 22 頁）

(21)出願番号　特願2014-48187(P2014-48187)
(22)出願日　平成26年3月11日(2014.3.11)
(31)優先権主張番号　特願2013-48473(P2013-48473)
(32)優先日　平成25年3月11日(2013.3.11)
(33)優先権主張国　日本国(JP)

特許法第３０条第２項適用申請有り

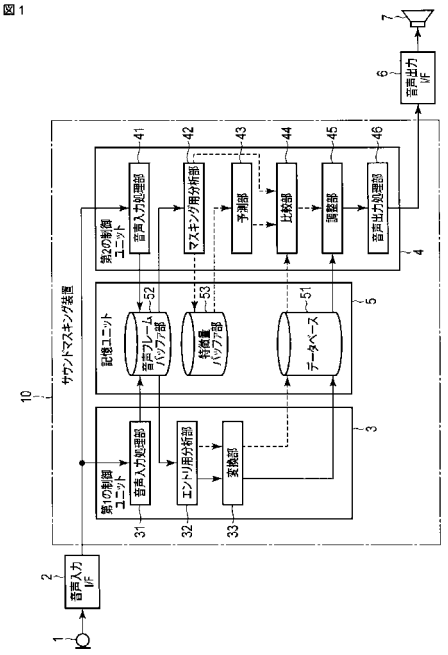
(71)出願人　502350504
学校法人上智学院
東京都千代田区紀尾井町７番１号
(74)代理人　100108855
弁理士　蔵田　昌俊
(74)代理人　100109830
弁理士　福原　淑弘
(74)代理人　100103034
弁理士　野河　信久
(74)代理人　100075672
弁理士　峰　隆司
(74)代理人　100153051
弁理士　河野　直樹
(74)代理人　100140176
弁理士　砂川　克

最終頁に続く

(54)【発明の名称】サウンドマスキング装置、方法及びプログラム

(57)【要約】
【課題】マスキング対象の音声が変化した場合でも、音響特性の条件を満たすマスキング音を生成して、特定ユーザの音が変化した場合や不特定ユーザに対しても高いマスキング効果が得られるようにする。
【解決手段】データベース５１にエントリデータを登録するための第１の制御ユニット３に、音声入力処理部３１及び分析部３２に加え変換部３３を設けている。そしてこの変換部３３において、分析部３２により入力音声データの音声フレーム毎に算出された音声パラメータのフレーム内平均値を、複数段階に変化させ、この変化後の各パラメータ値に対応するように上記音声フレームデータを変換し、この変換された音声フレームデータと対応する音声パラメータ値との対をデータベース５１に格納するようにしたものである。
【選択図】図１

図1



【特許請求の範囲】
【請求項１】
複数のマスキング用の音データがその音響特性を表す情報と共に格納されたデータベースとの間でデータ伝送が可能であり、マスキング対象の音データの音響特性との間で予め設定した関係条件を満たすマスキング用音データを前記データベースから選択して当該音データに対応する音を出力するサウンドマスキング装置であって、
前記マスキング用音データを生成して前記データベースに格納するための第１の制御ユニットを具備し、
前記第１の制御ユニットは、
標本音声の入力を受付け、この入力された標本音声を予め定めたフレーム長で複数のフレームに分割する第１の分割手段と、
前記分割されたフレーム毎に当該音データの音響特性を分析して当該音響特性を表すパラメータ値を計算する第１の分析手段と、
前記第１の分析手段により計算されたパラメータ値を予め設定した間隔で段階的に変化させて異なる複数の新たなパラメータ値を算出し、前記フレームの音データを、そのパラメータ値が前記算出された複数の新たなパラメータ値となるように変換する変換手段と、
前記変換手段により変換された複数の音データを前記マスキング用音データとして、当該音データと対応するパラメータ値と関連付けて前記データベースに格納する記憶制御手段と
を具備するサウンドマスキング装置。
【請求項２】
前記データベースからマスキング用音データを選択して当該音データに対応する音を出力する第２の制御ユニットを、さらに具備し、
前記第２の制御ユニットは、
マスキング対象の音データの入力を受付け、この入力された音データを前記フレーム長で複数のフレームに分割する第２の分割手段と、
前記分割されたフレーム毎に当該マスキング対象の音データの音響特性を分析して当該音響特性を表すパラメータ値を計算する第２の分析手段と、
前記第２の分析手段により計算されたマスキング対象の音データのパラメータ値を前記データベースに格納されている複数のパラメータ値と比較し、前記計算されたマスキング対象の音データとの間でパラメータ値が予め設定した関係条件を満たすマスキング用音データを選択する比較手段と、
前記選択されたマスキング用音データを、その対応するパラメータ値と前記計算されたマスキング対象の音データのパラメータ値との関係が前記関係条件の中の最良の条件を満たすように調整する調整手段と、
前記調整手段により調整されたマスキング用音データに対応する音をスピーカから出力させる手段と
を備えることを特徴とする請求項１記載のサウンドマスキング装置。
【請求項３】
前記第２の制御ユニットは、
前記第２の分析手段により計算されたマスキング対象の音データのパラメータ値をバッファに保存させる手段と、
前記第２の分析手段により計算された現フレームにおけるマスキング対象音データのパラメータ値と、前記バッファに保存されている過去のフレームにおけるマスキング対象音データのパラメータ値をもとに、後続フレームにおけるパラメータ値を予測し、この予測されたパラメータ値を、前記第２の分析手段により計算された現フレームにおけるマスキング対象音データのパラメータ値に代えて前記比較手段に与える予測手段と
を、さらに具備することを特徴とする請求項２記載のサウンドマスキング装置。
【請求項４】

前記第１の分割手段は、標本音声として日本語の単音節音声を複数個選択し、それぞれの単音節音声からそのフォルマント遷移部を中心に子音部の末尾部分と母音部の冒頭部分とを含む１フレーム長の音データを抽出し、
前記第１の分析手段は、
前記単音節音声毎に、前記抽出された１フレーム長の音データの音響特性を分析して当該音響特性を表すパラメータ値を算出する手段と、
前記算出されたパラメータ値をもとに、当該パラメータ値と関連する複数の新たなパラメータ値を算出する手段と
を備え、
前記変換手段は、前記単音節音声毎に抽出された１フレーム長の音データを、そのパラメータ値が前記算出された複数の新たなパラメータ値となるようにそれぞれ変換し、
前記記憶制御手段は、前記単音節音声毎に抽出された１フレーム長の音データを親エントリとし、かつ前記変換された複数の音データを子エントリとして、これらの音データを前記マスキング用音データとして、当該音データと対応する各パラメータ値と関連付けて前記データベースに格納する
ことを特徴とする請求項１記載のサウンドマスキング装置。
【請求項５】
前記第１の制御ユニットは、
基本周波数が第１の周波数帯域に含まれる標本音声に基づいて第１のマスキング用音データを生成して、当該第１のマスキング用音データを当該音データと対応するパラメータ値に関連付けて前記データベースに格納する処理機能と、
基本周波数が前記第１の周波数帯域とは異なる第２の周波数帯域に含まれる標本音声に基づいて第２のマスキング用音データを生成して、当該第２のマスキング用音データを当該音データと対応するパラメータ値に関連付けて前記データベースに格納する処理機能と
を備え、
前記第２の制御ユニットは、マスキング対象の音データが入力された場合に、当該入力された音データの音響特性を表すパラメータ値に含まれる基本周波数をもとに、前記データベースから前記第１及び第２のマスキング用音データの一方を選択的に読み出すことを特徴とする請求項２記載のサウンドマスキング装置。
【請求項６】
前記第１又は第２の分析手段は、前記パラメータ値として音データの基本周波数を計算することを特徴とする請求項１乃至５のいずれかに記載のサウンドマスキング装置。
【請求項７】
複数のマスキング用の音データがその音響特性を表す情報と共に格納されたデータベースとの間でデータ伝送が可能であり、マスキング対象の音データの音響特性との間で予め設定した関係条件を満たすマスキング用音データを前記データベースから選択して当該音データに対応する音を出力するサウンドマスキング装置であって、
マスキング対象の音データの入力を受付け、この入力された標本音声を前記フレーム長で複数のフレームに分割する手段と、
前記分割されたフレーム毎に当該マスキング対象の音データの音響特性を分析して当該音響特性を表すパラメータ値を計算する分析手段と、
前記分析手段により計算されたマスキング対象音データのパラメータ値を前記データベースに格納されている複数のパラメータ値と比較し、前記計算されたマスキング対象の音データとの間でパラメータ値が予め設定した関係条件を満たすマスキング用音データを選択する比較手段と、
前記選択されたマスキング用音データを、その対応するパラメータ値と前記計算されたマスキング対象の音データのパラメータ値との関係が前記関係条件の中の最良の条件を持たすように調整する調整手段と、
前記調整手段により調整されたマスキング用音データに対応する音をスピーカから出力

させる手段と
を具備することを特徴とするサウンドマスキング装置。
【請求項８】
前記分析手段により計算されたマスキング対象の音データのパラメータ値をバッファに保存させる手段と、
前記分析手段により計算された現フレームにおけるマスキング対象音データのパラメータ値と、前記バッファに保存されている過去のフレームにおけるマスキング対象音データのパラメータ値をもとに、後続フレームにおけるパラメータ値を予測し、この予測されたパラメータ値を、前記分析手段により計算された現フレームにおけるマスキング対象音データのパラメータ値に代えて前記比較手段に与える予測手段と
を、さらに具備することを特徴とする請求項７記載のサウンドマスキング装置。
【請求項９】
複数のマスキング用の音データがその音響特性を表す情報と共に格納されたデータベースとの間でデータ伝送が可能であり、マスキング対象の音データの音響特性との間で予め設定した関係条件を満たすマスキング用音データを前記データベースから選択して当該音データに対応する音を出力するサウンドマスキング装置が実行するデータベース作成方法であって、
標本音声の入力を受付け、この入力された標本音声を予め定めたフレーム長で複数のフレームに分割する過程と、
前記分割されたフレーム毎に当該音データの音響特性を分析して当該音響特性を表すパラメータ値を計算する過程と、
前記計算されたパラメータ値を予め設定した間隔で段階的に変化させて異なる複数のパラメータ値を算出し、前記フレームの音データを、そのパラメータ値が前記算出された複数のパラメータ値となるように変換する過程と、
前記変換された複数の音データを前記マスキング用音データとして、当該音データと対応するパラメータ値と関連付けて前記データベースに格納する過程と
を具備するサウンドマスキング方法。
【請求項１０】
マスキング対象の音データの入力を受付け、この入力された音データを前記フレーム長で複数のフレームに分割する過程と、
前記分割されたフレーム毎に当該マスキング対象の音データの音響特性を分析して当該音響特性を表すパラメータ値を計算する過程と、
前記計算されたマスキング対象音データのパラメータ値を前記データベースに格納されている複数のパラメータ値と比較処理し、前記計算されたマスキング対象の音データとの間でパラメータ値が予め設定した関係条件を満たすマスキング用音データを選択する過程と、
前記選択されたマスキング用音データを、その対応するパラメータ値と前記計算されたマスキング対象の音データのパラメータ値との関係が前記関係条件の中の最良の条件を満たすように調整する過程と、
前記調整されたマスキング用音データに対応する音をスピーカから出力させる過程と
を、さらに具備することを特徴とする請求項９記載のサウンドマスキング方法。
【請求項１１】
前記計算されたマスキング対象の音データのパラメータ値をバッファに保存させる過程と、
前記計算された現フレームにおけるマスキング対象音データのパラメータ値と、前記バッファに保存されている過去のフレームにおけるマスキング対象音データのパラメータ値をもとに、後続フレームにおけるパラメータ値を予測し、この予測されたパラメータ値を、前記計算された現フレームにおけるマスキング対象音データのパラメータ値に代えて前記比較処理に供する過程と
を、さらに具備することを特徴とする請求項１０記載のサウンドマスキング方法。

【請求項１２】
複数のマスキング用の音データがその音響特性を表す情報と共に格納されたデータベースとの間でデータ伝送が可能であり、マスキング対象の音データの音響特性との間で予め設定した関係条件を満たすマスキング用音データを前記データベースから選択して当該音データに対応する音を出力するサウンドマスキング装置が実行するサウンドマスキング方法であって、
マスキング対象の音データの入力を受付け、この入力された音データを前記フレーム長で複数のフレームに分割する過程と、
前記分割されたフレーム毎に当該マスキング対象の音データの音響特性を分析して当該音響特性を表すパラメータ値を計算する過程と、
前記計算されたマスキング対象音データのパラメータ値を前記データベースに格納されている複数のパラメータ値と比較し、前記計算されたマスキング対象の音データとの間でパラメータ値が予め設定した関係条件を満たすマスキング用音データを選択する過程と、
前記選択されたマスキング用音データを、その対応するパラメータ値と前記計算されたマスキング対象の音データのパラメータ値との関係が前記関係条件の中の最良の条件を持たすように調整する過程と、
前記調整されたマスキング用音データに対応する音をスピーカから出力させる過程と
を具備することを特徴とするサウンドマスキング方法。
【請求項１３】
請求項１乃至８のいずれかに記載のサウンドマスキング装置が具備する各手段による処理を、当該サウンドマスキング装置が備えるコンピュータに実行させるプログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
この発明は、スピーチプライバシを保護する手法の１つとして用いられるサウンドマスキング装置と、このサウンドマスキング装置が実行するサウンドマスキング方法及びプログラムに関する。
【背景技術】
【０００２】
対象音が聞こえている状態で同一空間に当該対象音に近い音響特性を持つ別の音が存在すると対象音が聞こえにくくなるという現象が一般に知られている。この現象はマスキング効果と呼ばれ、別の音として使用されるマスキング音の周波数が対象音の周波数に近いほど、またマスキング音の音量レベルが対象音の音量レベルに対して相対的に高いほど顕著になる。
【０００３】
そこで、このマスキング効果を利用し、役所や病院、薬局の受付等において話者のスピーチプライバシを保護するために、話者の話し声をマスキングして周囲にいる第三者に聞かれないようにするサウンドマスキングシステムが種々提案されている。例えば、特許文献１には、先ず言語としての意味が判別できないように処理されたスクランブル音信号をその音響特性を表す情報と共に事前にテーブルに格納しておき、音響空間における音を表す音信号を受け取ると、この音信号の音響特性を分析し、当該分析された音響特性と特性が最も類似するスクランブル音信号を上記テーブルから選択してその音を音響空間へ出力する技術が記載されている。
【先行技術文献】
【特許文献】
【０００４】
【特許文献１】特開２００８－２３３６７２号公報
【発明の概要】
【発明が解決しようとする課題】
【０００５】

ところが、特許文献１に記載された技術では、マスキング対象となるユーザが入力した音声信号をスクランブル処理した音信号と、このスクランブル音信号から抽出した音響特性情報をただ単にテーブルに格納するようにしている。このため、不特定ユーザの音声をマスキングしようとすると、ユーザが持つ声の癖等によっては期待するマスキング効果が得られない。また、不特定ユーザに対し漏れなく高いマスキング効果を得るためには、マスキング対象のユーザが変わるごとにデータベースを構築し直さなければならず、その処理負荷がきわめて大きくなる。また、公共の場所に適用することができない。

【０００６】

この発明は上記事情に着目してなされたもので、その目的とするところは、マスキング対象の音声が変化した場合でも、音響特性の条件を満たすマスキング音を出力できるようにし、これにより特定ユーザの音が変化した場合や不特定ユーザに対しても高いマスキング効果が得られるようにしたサウンドマスキング装置、方法及びプログラムを提供することにある。

【課題を解決するための手段】

【０００７】

上記目的を達成するためにこの発明の第１の観点は、複数のマスキング用の音データをその音響特性を表す情報と共にデータベースに格納しておき、マスキング対象の音データの音響特性との間で予め設定した関係条件を満たすマスキング用の音データを上記データベースから選択して当該音データに対応する音を出力するサウンドマスキング装置にあって、先ず標本音声の入力を受付けて、この入力された標本音声を予め定めたフレーム長で複数のフレームに分割する。次に、上記分割されたフレーム毎に当該音データの音響特性を分析して当該音響特性を表すパラメータ値を計算し、この計算されたパラメータ値を予め設定した間隔で段階的に変化させて異なる複数の新たなパラメータ値を算出し、上記フレームの音データを、そのパラメータ値が上記算出された複数の新たなパラメータ値となるように変換する。そして、この変換された複数の音データをマスキング用音データとして、当該音データと対応するパラメータ値と共に上記データベースに格納するようにしたものである。

【０００８】

この発明の第２の観点は、上記第１の観点に加えて、さらに以下の処理を行うようにしたものである。すなわち、マスキング対象の音データの入力を受付けると、先ずこの入力された音データを前記フレーム長で複数のフレームに分割して、この分割されたフレーム毎に当該マスキング対象の音データの音響特性を分析して当該音響特性を表すパラメータ値を計算する。次に、この計算されたマスキング対象音データのパラメータ値を前記データベースに格納されている複数のパラメータ値と比較し、前記計算されたマスキング対象の音データとの間でパラメータ値が予め設定した関係条件を満たすマスキング用音データを選択する。さらに、上記選択されたマスキング用音データを、その対応するパラメータ値と前記計算されたマスキング対象の音データのパラメータ値との関係が前記関係条件の中の最良の条件を満たすように調整し、この調整されたマスキング用音データに対応する音をスピーカから出力させるようにしたものである。

【０００９】

この発明の第３の観点は、上記第２の観点に加えて、さらに以下の処理機能を備えるようにしたものである。すなわち、上記計算されたマスキング対象の音データのパラメータ値をバッファに保存しておき、上記計算された現フレームにおけるマスキング対象音データのパラメータ値と、上記バッファに保存されている過去のフレームにおけるマスキング対象音データのパラメータ値をもとに、後続フレームにおけるパラメータ値を予測する。そして、この予測されたパラメータ値を、上記計算された現フレームにおけるマスキング対象音データのパラメータ値に代えて上記比較処理に供するようにしたものである。

【００１０】

この発明の第４の観点は、上記第１の制御ユニットが以下のような処理機能を備えたものである。すなわち、先ず前記第１の分割手段により、標本音声として日本語の単音節音

声を複数個選択し、それぞれの単音節音声からそのフォルマント遷移部を中心に子音部の末尾部分と母音部の冒頭部分とを含む１フレーム長の音データを抽出する。次に、前記第１の分析手段により、前記単音節音声毎に、前記抽出された１フレーム長の音データの音響特性を分析して当該音響特性を表すパラメータ値を算出し、前記算出されたパラメータ値をもとに当該パラメータ値と関連する複数の新たなパラメータ値を算出する。そして、前記変換手段により、前記単音節音声毎に抽出された１フレーム長の音データを、そのパラメータ値が前記算出された複数の新たなパラメータ値となるようにそれぞれ変換し、前記記憶制御手段により、前記単音節音声毎に抽出された１フレーム長の音データを親エントリとし、かつ前記変換された複数の音データを子エントリとして、これらの音データを前記マスキング用音データとして、当該音データと対応する各パラメータ値と関連付けて前記データベースに格納するようにしたものである。

【００１１】

この発明の第５の観点は、前記第１の制御ユニットにより、基本周波数が第１の周波数帯域に含まれる標本音声に基づいて第１のマスキング用音データを生成して、当該第１のマスキング用音データを当該音データと対応するパラメータ値に関連付けて前記データベースに格納すると共に、基本周波数が前記第１の周波数帯域とは異なる第２の周波数帯域に含まれる標本音声に基づいて第２のマスキング用音データを生成して、当該第２のマスキング用音データを当該音データと対応するパラメータ値に関連付けて前記データベースに格納する。そして、マスキング対象の音データが入力された場合に、前記第２の制御ユニットにより、当該入力された音データの音響特性を表すパラメータ値に含まれる基本周波数をもとに、前記データベースから前記第１及び第２のマスキング用音データの一方を選択的に読み出すようにしたものである。

【００１２】

この発明の第６の観点は、上記標本音声又はマスキング対象の音声の音響特性を表す情報として、音データの基本周波数を計算するものである。

【発明の効果】

【００１３】

この発明の第１の観点によれば、データベースには、標本音声データをもとにその音パラメータ値を段階的に変化させることによりそれぞれ作成された複数の変換音データが格納されることになる。このため、不特定多数のユーザの音声がマスキング対象として入力された場合でも、当該入力された音声データとの間で音声パラメータ値が予め設定した関係条件を満たすマスキング用音データをデータベースから選択できる確率が高くなり、これにより上記マスキング対象の音声を効果的にマスキングすることが可能となる。

【００１４】

この発明の第２の観点によれば、マスキング対象音声の音パラメータ値との間で最良の関係条件を満たすマスキング用音データをデータベースから選択できなかった場合でも、当該マスキング用音データが上記最良の関係条件を満たすように調整されるため、マスキング対象の音声をさらに効果的にマスキングすることが可能となる。また、上記のようにマスキング用音データが調整されることにより、データベースへのマスキング用音データのエントリ数を減らすことができ、これによりデータベースの記憶容量を削減すると共に、データベースからマスキング用音データを選択する際のアクセス所用時間を短縮して、マスキング音の出力遅延を減少させることが可能となる。

【００１５】

この発明の第３の観点によれば、マスキング音の遅延量をさらに減らす必要がある場合に、過去に予測しておいたパラメータ値をもとにデータベースからマスキング音データが選択されるので、マスキング対象音声の分析処理等による遅延が発生する場合でも、高いマスキング効果を得ることが可能となる。

【００１６】

この発明の第４の観点によれば、日本語の単音節音声、例えば清音、濁音、半濁音、拗音の各々からそのフォルマント遷移部を中心に子音部の末尾部分と母音部の冒頭部分とを

含む１フレーム長の音データが抽出され、この単音節毎に抽出された音声フレームが親エントリとして、また上記単音節毎に抽出された音声フレームから変換された複数の新たな音声データが子エントリとしてデータベースに格納される。すなわち、単音節ごとにフォルマント遷移部を中心に子音部の末尾部分と母音部の冒頭部分とを含む領域のみについて、マスキング用音データのエントリデータ群が生成されてデータベースに格納される。

【００１７】

したがって、単音節毎にその全領域を複数のフレームに分割し、これらのフレーム毎にマスキング用音データのエントリデータ群生成してデータベースにエントリする場合に比べ、データベースへのエントリデータ量とエントリに必要な処理時間を大幅に減らすことができ、さらにデータベースからマスキング対象の音データに適したマスキング用音データを選択するために必要な時間を短縮して、マスキング処理の応答性を高めることができる。

【００１８】

この発明の第５の観点によれば、周波数特性が互いに異なる２つの標本音声をもとにそれぞれマスキング用音データのエントリデータ群が生成されてデータベースに格納され、マスキング対象の音データの基本周波数に応じて上記データベースから当該マスキング対象の音データに適したマスキング用音データがデータベースから選択されその音が出力される。このため、例えば男性と女性に対しそれぞれ適切なマスキング用音データを自動的に選択してマスキングを行うことが可能となる。

【００１９】

この発明の第６の観点によれば、入力音声の音パラメータとして入力音の基本周波数が算出され、この基本周波数をもとにマスキング用音データの作成処理、及びマスキング対象音声と類似するマスキング音の生成処理が行われる。このため、音パラメータとしてスペクトル包絡等を用いる場合に比べ、高いマスキング効果が期待できる。

【００２０】

すなわちこの発明によれば、マスキング対象の音声が変化した場合でも、音響特性の条件を満たすマスキング音を出力できるようにし、これにより特定ユーザの音が変化した場合や不特定ユーザに対しても高いマスキング効果が得られるようにしたサウンドマスキング装置、方法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】

【００２１】

【図１】この発明の第１の実施形態に係るサウンドマスキング装置の機能構成を示すブロック図。

【図２】図１に示したサウンドマスキング装置のエントリ用分析処理及び変換処理の手順と処理内容を示すフローチャート。

【図３】図１に示したサウンドマスキング装置のマスキング用分析処理、予測処理、比較処理及び調整処理の手順と処理内容を示すフローチャート。

【図４】図１に示したサウンドマスキング装置によるマスキング効果の第１の例を説明するための図。

【図５】図１に示したサウンドマスキング装置によるマスキング効果の第２の例を説明するための図。

【図６】図１に示したサウンドマスキング装置による予測処理の概要を説明するための図。

【図７】図１に示したサウンドマスキング装置による予測処理の効果を説明するための図。

【図８】この発明の第２の実施形態に係るサウンドマスキング装置の要部の機能構成を示すブロック図。

【図９】この発明の第３の実施形態に係るサウンドマスキング装置で使用されるMiddleデータベースの効果を説明するための図。

【発明を実施するための形態】

【００２２】
以下、図面を参照してこの発明に係わる実施形態を説明する。
（構成）
図１は、この発明の第１の実施形態に係るサウンドマスキング装置の機能構成を示すブロック図であり、図中１０がサウンドマスキング装置を示している。
サウンドマスキング装置１０には、音響空間に設置されたマイクロホン１及びスピーカ７がそれぞれ音声入力インタフェース（音声入力Ｉ／Ｆ）２及び音声出力インタフェース（音声出力Ｉ／Ｆ）６を介して接続されている。音声入力Ｉ／Ｆ２は、マイクロホン１から出力されたアナログ音声信号をディジタル音声信号に変換してサウンドマスキング装置１０に入力する機能を有する。音声出力Ｉ／Ｆ６は、サウンドマスキング装置１０から出力されたマスキング用の音データをアナログのマスキング音信号に変換したのち、増幅してスピーカ７から拡声出力させる機能を有する。
【００２３】
サウンドマスキング装置１０は、例えばパーソナル・コンピュータからなり、第１の制御ユニット３と、第２の制御ユニット４と、記憶ユニット５を備えている。
記憶ユニット５は、記憶媒体としてＨＤＤ（Hard Disk Drive）又はＳＳＤ（Solid State Drive）を備え、この実施形態を実施する上で必要な記憶領域として、データベース５１と、音声フレームバッファ部５２と、特徴量バッファ部５３を有している。
【００２４】
データベース５１は、後述する第１の制御ユニット３により作成されたマスキング用の音データとその音響特性を表す音声パラメータとからなるエントリを、複数個格納するために用いられる。音声フレームバッファ部５２は、フレーム化された入力音声データを一時保存するために使用される。特徴量バッファ部５３は、音声フレーム毎に分析され得られた音声パラメータの特徴量を表すデータを、後述する予測部４３による特徴量予測処理のために保存する。
【００２５】
第１及び第２の制御ユニット３，４はいずれもＣＰＵ（Central Processing Unit）及びＤＳＰ（Digital Signal Processor）を備える。なお、これらのＣＰＵ及びＤＳＰは、第１の制御ユニット３と第２の制御ユニット４に対し共通に設けてもよく、また別々に設けてもよい。
【００２６】
第１の制御ユニット３は、データベース５１に格納するエントリ群を作成するためのもので、音声入力処理部３１と、エントリ用分析部３２と、変換部３３を有している。なお、図１中の実線の矢印は音声データの流れを示し、また破線の矢印は音声パラメータの流れを示す。
【００２７】
音声入力処理部３１は、上記音声入力Ｉ／Ｆ２から標本用のディジタル音声信号を受け取り、この受け取ったディジタル入力音声信号を予め設定された時間長で複数の音声フレームに分割して、上記音声フレームバッファ部５２に保存させる。１フレーム長は例えば１００msに設定されるが、その他の長さに設定してもよい。
【００２８】
エントリ用分析部３２は、上記音声フレームバッファ部５２から入力音声データを１フレームずつ読み込み、この読み込んだ音声フレームから音声パラメータを抽出する計算を行う。抽出対象となる音声パラメータには、例えば基本周波数Ｆ０と、スペクトル特性と、強度（例えば音量レベル）がある。エントリ用分析部３２はさらに、上記音声フレーム毎に抽出されたパラメータについてそれぞれフレーム内の平均値を算出する。
【００２９】
変換部３３は、上記算出された各音声パラメータのフレーム内平均値をそれぞれ段階的に変化させ、この変化後のパラメータ値に対応するように上記音声フレームのデータを変換する。そして、この変換後の音声フレームデータとこれに対応する上記変化後のパラメ

ータ値との対を１つのエントリデータとしてデータベース５１に格納する処理を行う。
【００３０】
第２の制御ユニット４は、マスキング対象となる音声が入力された場合にマスキング用の音データを生成するもので、音声入力処理部４１と、マスキング用分析部４２と、予測部４３と、比較部４４と、調整部４５と、音声出力処理部４６を有している。なお、ここでも図中の実線の矢印は音声データの流れを示し、また破線の矢印は音声パラメータの流れを示す。
【００３１】
音声入力処理部４１は、上記音声入力Ｉ／Ｆ２からマスキング対象のディジタル音声信号を受け取り、この受け取ったディジタル入力音声信号を上記標本用のディジタル音声信号のフレーム長と同一のフレーム長で分割して、上記音声フレームバッファ部５２に保存させる。
【００３２】
マスキング用分析部４２は、上記音声フレームバッファ部５２からマスキング対象のディジタル音声データを１フレームずつ読み込み、この読み込んだ音声フレームから音声パラメータを抽出する計算を行う。抽出対象となる音声パラメータは、先に述べたエントリ用分析部３２と同様に、基本周波数Ｆ０と、スペクトル特性と、強度からなる。エントリ用分析部３２はさらに、上記音声フレーム毎に抽出されたパラメータについてそれぞれフレーム内の平均値を算出し、この算出された各パラメータのフレーム内平均値を特徴量バッファ部５３に一時保存させる処理を行う。
【００３３】
予測部４３は、上記マスキング用分析部４２により算出された現フレームの音声パラメータ値と、上記特徴量バッファ部５３に記憶された過去の複数のフレームの音声パラメータ値をもとに、数フレーム先の音声パラメータのフレーム内平均値を予測する処理を行う。
【００３４】
比較部４４は、上記マスキング用分析部４２により算出された現フレームの音声パラメータのフレーム内平均値と、上記予測部４３により予測された各音声パラメータのフレーム内平均値とのいずれか一方を、データベース５１に格納されている各エントリデータの音声パラメータのフレーム内平均値と順次比較する。そして、データベース５１に格納されている各エントリデータの中で、上記入力音声データから算出した音声パラメータのフレーム内平均値、或いはその予測値に対し、音声パラメータの条件を満たすエントリデータを選択する処理を行う。
【００３５】
なお、現フレームの音声パラメータのフレーム内平均値と、予測された音声パラメータのフレーム内平均値とのいずれを使用するかは、装置の管理者が手動で設定する。また他の選択手法として、例えばマスキング対象となる音データの音量レベルに応じて自動的に選択するようにしてもよい。例えば、当該音量レベルが閾値以上の場合には、マスキング対象音声に対しパラメータ値がより近いマスキング用音声を使用する必要があると考えられるため、予測された音声パラメータのフレーム内平均値を選択する。これに対しマスキング対象の音声の音量レベルが閾値未満の場合には、マスキング対象音声に対しパラメータ値がそれほど近くなくても一定のマスキング効果が得られると考えられるので、この場合には現フレームの音声パラメータのフレーム内平均値を選択する。また、マスキング対象の音声の音響特性によらず、常に、予測された音声パラメータのフレーム内平均値を選択するようにしてもよい。
【００３６】
調整部４５は、上記比較部４４により選択されたエントリデータの音声フレームを、当該エントリデータの音声パラメータ値が上記マスキング用分析部４２により算出された現フレームの音声パラメータ値と一致するように調整する処理を行う。
【００３７】

音声出力処理部４６は、上記調整部４５により調整された音声フレームを接続して連続する音声データを生成し、この生成された音声データを音声出力Ｉ／Ｆ６へ出力する処理を行う。

【００３８】

なお、上記第１及び第２の制御ユニット３，４が備える各制御機能は、何れも図示しないプログラム・メモリに格納されたアプリケーション・プログラムを上記ＣＰＵ又はＤＳＰに実行させることにより実現される。

【００３９】

（動作）

次に、以上のように構成されたサウンドマスキング装置１０の動作を説明する。

（１）データベースの作成

先ず、標本として任意に選んだ人が発声を開始し、その音声がマイクロホン１に入力されると、この入力音声に対応する音声信号がマイクロホン１から出力され、音声入力Ｉ／Ｆ２でディジタル信号に変換されたのちサウンドマスキング装置１０に入力される。なお、上記標本となる音声を発する人は一人でもよいが複数でもよい。

【００４０】

サウンドマスキング装置１０では、上記入力されたディジタル音声信号が第１の制御ユニット３の音声入力処理部３１に所定のフレーム長（例えば１００ms）ずつ取り込まれ、この取り込まれた音声フレームが時系列に従い音声フレームバッファ部５２に一時保存される。すなわち、この処理により入力ディジタル音声信号は１００msのフレーム長に分割される。

【００４１】

なお、音声フレームの長さは１００ms以外に設定してもよく、さらに要求されるマスキング効果の高さや遅延量に応じて可変設定するようにしてもよい。また、上記入力された一定長分のディジタル音声信号を一旦バッファメモリに蓄積し、しかるのち当該ディジタル音声信号を読み出して一定フレーム長に分割するようにしてもよい。

【００４２】

次に第１の制御ユニット３では、エントリ用分析部３２及び変換部３３により、音声フレームに対し以下のような分析処理及び変換処理が実行される。図２はその処理手順と処理内容を示すフローチャートである。

すなわち、先ずステップＳ１１において、エントリ用分析部３２の制御の下、上記音声フレームバッファ部５２から入力音声データＳが１フレームずつ読み込まれ、この読み込まれた音声フレームから音声パラメータＰｉ を抽出する計算が行われる。なお、ここでは音声パラメータＰｉ として、例えば基本周波数Ｆ０と、スペクトル特性と、音量レベルが抽出される。そして、この抽出された３種類の音声パラメータＰｉ （３種類なのでi=1,2,3）についてそれぞれフレーム内平均値が算出される。

【００４３】

次に変換部３３の制御の下で、上記算出された各音声パラメータＰｉ のフレーム内平均値をそれぞれ複数段階に変化させ、この変化後の各パラメータ値に対応するように上記音声フレームデータＳを変換する処理が行われる。

すなわち、段階数がｍ（ｍはインデックスで整数値（ｍ＝－Ｍ～Ｍ））であるとき、先ずステップＳ１２においてｍが初期値－Ｍに設定される。次にステップＳ１３において、上記段階－Ｍにおける音声パラメータＰｉ,ｍ が

$$Pi,m = Pi + m \times \Delta Pi$$

として計算される。なお、ΔＰｉ は音声パラメータＰｉ を段階的に変化させるときのステップ幅である。

【００４４】

次にステップＳ１４において、上記音声フレームデータＳが、その音声パラメータＰｉ が上記計算された段階－Ｍにおける音声パラメータＰｉ,ｍ となるように変換される。そして、ステップＳ１５において、上記変換された音声フレームデータＳｉ,ｍと上記音声パラ

メータＰi,mとの対が１個のエントリデータとしてデータベース５１に格納される。

【００４５】

続いてステップＳ１６により、段階数がｍ＝Ｍに達したか否かが判定される。そして、ｍ＝Ｍに達していなければ、ステップＳ１７によりｍの値がインクリメント（ｍ＝ｍ＋１）された後、ステップＳ１３に戻って上記ステップＳ１３～Ｓ１５による音声フレームデータの変換処理及びデータベース５１へのエントリデータの登録処理が行われる。以後同様に、ｍ＝Ｍに達するまで各段階数ｍにおける上記ステップＳ１３～Ｓ１５による音声フレームデータの変換処理及びデータベース５１へのエントリデータの登録処理が繰り返し実行される。

【００４６】

例えば、段階数ｍとして、上記算出された音声フレームの基本周波数Ｆ０の平均値に対し±２５Hz、±５０Hz、±７５Hzの６段階を設定したとする。この場合、先ず上記音声フレームデータＳが、その基本周波数Ｆ０が上記－７５Hzのときの音声フレームデータに変換される。そして、この変換された音声フレームデータとＦ０－７５Hzの周波数値との対が１個のエントリデータとしてデータベース５１に格納される。次に、上記音声フレームデータＳが、その基本周波数Ｆ０が上記－５０Hzのときの音声フレームデータに変換され、Ｆ０－５０Hzの周波数値と共にデータベース５１に格納される。同様に、上記音声フレームデータＳが、その基本周波数Ｆ０が上記－２５Hz、＋２５Hz、＋５０Hz、＋７５Hzのときの音声フレームデータにそれぞれ変換され、対応する周波数値と共にデータベース５１に格納される。

【００４７】

以下同様に、スペクトル特性及び強度（例えば音量レベル）についても、それぞれｍ段階に変化させたときのパラメータ値となるように入力音声フレームデータが変換され、この変換された音声フレームデータが対応する変化後のパラメータ値と共にデータベース５１に格納される。

【００４８】

かくして、データベース５１には、段階数ｍの１段階ごとに、入力音声フレームデータの音声パラメータＰi の変化後の値の全ての組み合わせについてそれぞれ変換された音声フレームデータＳi,mがそれぞれエントリデータとして登録される。

【００４９】

例えば、音声パラメータＰi が先に述べた３種類（i=1,2,3）の場合であれば、ｍ（－Ｍ～Ｍ）の各段階ごとに、基本周波数Ｆ０、フォルマント及び音量レベルをそれぞれ単独で変化させたときの変換後の音声フレームデータと、基本周波数Ｆ０とフォルマントを変化させたときの変換後の音声フレームデータと、基本周波数Ｆ０と音量レベルを変化させたときの変換後の音声フレームデータと、フォルマントと音量レベルを変化させたときの変換後の音声フレームデータと、基本周波数Ｆ０、フォルマント及び音量レベルを全て同時に変化させたときの変換後の音声フレームデータとからなる、合計７個のエントリデータが登録される。そして、段階数ｍが６であれば、７×６＝４２個のエントリデータが登録される。なお、ｍを変化させないときの変換前の音声フレームデータもエントリデータの１つとして登録される。

【００５０】

なお、以上述べたデータベース５１へのエントリデータの登録処理は、予め設定された時間長の入力音声データに対し行われ、当該時間長分の入力音声データに基づくエントリデータの登録処理が終了すると、登録処理は終了となる。

【００５１】

上記データベースの作成方法として、具体的には以下の手法が挙げられる。この手法はWholeデータベースを用いたもので、日本語の単音節音声（清音・濁音・半濁音・拗音）を複数個（例えば１００種類）選択し、それぞれの単音節音声の先頭からフレーム長間隔（例えば１００ms）で分割する。そして、この分割されたフレームを親エントリとしてデータベースに記憶させる。すなわち、１音節につき複数個（親エントリの個数は単音節音

声の長さや分割するフレーム長に依存）の親エントリが生成され、データベースに記憶される。

【００５２】

次に、上記親エントリのそれぞれについて所定の変換処理が行われて新たな複数の音声データのエントリが生成され、この新たな音声データのエントリ群が子エントリとしてデータベースに記憶される。なお、上記子エントリを生成するための変換処理としては、例えば基本周波数のピッチ変換が用いられる。ピッチ変換は、例えばそれぞれのフレームの平均基本周波数を操作（原音を－５０Hz，－４８Hz，…，－２Hz，＋２Hz，＋４Hz，…，＋１００Hz）することにより行う。なお、ピッチ変換を行う原音の周波数間隔は上記間隔に限定されるものではなく、任意に設定できる。また、上記子エントリを生成するための変換処理には、基本周波数のピッチ変換以外にスペクトル変換等を用いてもよい。

【００５３】

（２）オンラインにおけるマスキング用音データの生成処理

マスキング対象となるユーザが会話を開始し、その音声がマイクロホン１に入力されると、この入力音声に対応する音声信号がマイクロホン１から出力され、音声入力Ｉ／Ｆ２でディジタル信号に変換されたのちサウンドマスキング装置１０に入力される。

【００５４】

サウンドマスキング装置１０では、第２の制御ユニット４の音声入力処理部４１において、上記入力されたディジタル音声信号が前記第１の制御ユニット３において設定されたフレーム長（例えば１００ms）で分割され、この分割された音声フレームが時系列に従い音声フレームバッファ部５２に一時保存される。

【００５５】

次に第２の制御ユニット４では、マスキング用分析部４２、予測部４３、比較部４４、調整部４５及び音声出力処理部４６により、マスキング音データを生成するために以下のような処理が実行される。図３はその処理手順と処理内容を示すフローチャートである。

【００５６】

すなわち、先ずステップＳ２１において、上記音声フレームバッファ部５２から入力音声データSk が１フレームずつ読み込まれる。そしてステップＳ２２において、上記読み込まれた音声フレームデータから音声パラメータPinput を抽出する計算が行われる。なお、ここでも前記エントリ用分析部３２と同様に、音声パラメータPinput として、基本周波数Ｆ０、スペクトル特性及び音量レベルが抽出される。そして、この抽出された３種類の音声パラメータPi （３種類なのでi=1,2,3）についてそれぞれフレーム内平均値が算出される。なお、音声パラメータPinput としては、基本周波数Ｆ０、スペクトル特性及び音量レベルのうちのいずれか１つ又は２つを選択的に抽出するようにしてもよい。

【００５７】

また、ステップＳ２２において予測部４３では、上記マスキング用分析部４２から上記算出された現フレームの音声パラメータPinput のフレーム内平均値を受け取り、この現フレームの音声パラメータPinput のフレーム内平均値と、上記特徴量バッファ部５３に記憶されている過去の一定数分のフレームの音声パラメータのフレーム内平均値とをもとに、数フレーム先の音声パラメータＰ＾inputのフレーム内平均値が予測される。

【００５８】

次にステップＳ２３において、比較部４４の制御の下、上記マスキング用分析部４２で算出された現フレームの音声パラメータPinput のフレーム内平均値、または上記予測部４３により予測された音声パラメータＰ＾input のフレーム内平均値が、データベース５１に格納されている各エントリデータの音声パラメータPi,m のフレーム内平均値と順次比較される。

【００５９】

そして、音声パラメータが例えば基本周波数Ｆ０の場合或いは音量レベルの場合には、データベース５１に格納されている各エントリデータの中で、上記現フレームの音声パラメータPinput のフレーム内平均値、又は上記予測された音声パラメータＰ＾input のフ

レーム内平均値と最も類似する音声パラメータＰｋ のフレーム内平均値が選択される。
一方、音声パラメータがスペクトル特性の場合には、データベース５１に格納されている各エントリデータの中で、上記現フレームの音声パラメータＰinput のフレーム内平均値、又は上記予測された音声パラメータＰ＾input のフレーム内平均値に対し値が適度に離れている音声パラメータＰｋ のフレーム内平均値が選択される。

【００６０】

ところで、上記予測部４３による予測処理は、例えば以下のように行われる。図６に予測部４３を使用してサウンドマスキングを行うときの概念を示す。すなわち、マスキング用分析部４２では、一定間隔（例えば２０ms）で音声フレームの特徴量（例えば基本周波数及びフォルマント周波数）が分析され、特徴量バッファ部５３に格納される。予測部４３では、特徴量バッファ部５３に格納された最新の一定数のサンプル（例えば５サンプル）を用いて外挿予測が行われ、この処理により得られた特徴量が、未来のマスキング対象音に対するマスキング用音データの選択に使用される。具体的には、音声入力処理部４１から音声出力処理部４６までの各処理により発生する処理遅延の合計に相当する時間経過後に入力されるマスキング対象音のマスキングのために用いられる。

【００６１】

比較部４４では、上記予測処理により得られた未来のマスキング対象音に対しマスキングが最適に行われるようにするためのマスキング用音データが選択される。例えば、予測された音声パラメータが基本周波数であった場合、一般にマスキング対象の音声とマスキング音との間で基本周波数は近接していた方が好ましい。そこで、比較部４４では、予測部４３により予測された未来のマスキング対象音声の基本周波数に近い値を持ったパラメータ値が選択される。

【００６２】

続いてステップＳ２４において、調整部４５の制御の下で、上記選択された音声パラメータＰｋ のフレーム内平均値に対応する音声フレームデータＳｋ がデータベース５１から読み出される。そして、この読み出された音声フレームデータＳｋ が、その音声パラメータＰｋ のフレーム内平均値が上記現フレームの音声パラメータＰinput のフレーム内平均値、または上記予測された音声パラメータＰ＾input のフレーム内平均値と一致するように調整される。

【００６３】

最後にステップＳ２５において、音声出力処理部４６の制御の下、上記調整部４５により調整された音声フレームデータＳｋ が時系列に従い接続されて連続するディジタル音声信号が生成され、音声出力Ｉ／Ｆ６へ出力される。このディジタル音声信号は、音声出力Ｉ／Ｆ６によりアナログ音声信号に変換され、スピーカ７からマスキング音として拡声出力される。

かくして、マスキング対象のユーザの音声は上記スピーカ７から出力されるマスキング音によりマスキングされ、ユーザの音声のスピーチプライバシは保護される。

【００６４】

図４に、マスキング対象（ターゲット）の音声とマスキング音との音圧レベル比（ＴＭＲ；target-to-Masker Ratio）（dB）に対する単語了解度（％）の関係をロジスティック関数による回帰分析によって求めたものである。これによると、基本周波数Ｆ０をターゲットと類似させることで作成したマスキング音と、スペクトラム包絡をターゲットと類似させることで作成したＳＰＥＣマスキング音と、基本周波数及びスペクトラム包絡の何れも考慮せずにデータベース５１内のエントリデータを無作為に選択したＲＡＮＤＯＭマスキング音とを比較すると、基本周波数Ｆ０を類似させたマスキング音を発生させたときの単語了解度が最も低くなり、マスキング効果が最も高いことが確認できた。

【００６５】

また図５には、ＴＭＲと単語了解度との関係をロジスティック関数による回帰分析によって求めたものである。これによると、基本周波数Ｆ０をターゲットと類似させることにより作成したマスキング音と、基本周波数Ｆ０及びスペクトラム包絡の両方を考慮して作

成したＦ０＿ＳＰＥＣマスキング音と、白色雑音の低域が強調されたマスキング音Ｐｉｎｋとを比較すると、Ｐｉｎｋのマスキング音に比べ基本周波数Ｆ０を類似させたマスキング音、またはＦ０＿ＳＰＥＣマスキング音の方が単語了解度が低く抑えられ、マスキング効果が高いことが確認できた。

【００６６】

（効果）

以上詳述したようにこの発明の第１の実施形態では、データベース５１にエントリデータを登録するための第１の制御ユニット３に、音声入力処理部３１及び分析部３２に加え変換部３３を設けている。そしてこの変換部３３において、分析部３２により入力音声データの音声フレーム毎に算出された音声パラメータのフレーム内平均値を、複数段階に変化させ、この変化後の各パラメータ値に対応するように上記音声フレームデータを変換し、この変換された音声フレームデータと対応する音声パラメータ値との対をデータベース５１に格納するようにしている。

【００６７】

したがって、データベース５１には、任意ユーザの入力音声データをもとにその音声パラメータ値を段階的に変化させることによりそれぞれ作成された複数の変換音声データがエントリデータとして格納されることになる。このため、上記任意ユーザの声がマスキング対象として入力された場合でその音の高さ（ピッチ）等が変化した場合でも、また不特定多数のユーザの音声がマスキング対象として入力された場合でも、当該入力された音声データとの間で音声パラメータの関係条件を最も満足するエントリデータをデータベース５１から発見できる確率が高くなり、これにより上記マスキング対象の音声を効果的にマスキングすることが可能となる。

【００６８】

また本実施形態では、マスキング音を生成する第２の制御ユニット４に、音声入力処理部４１、マスキング用分析部４２及び比較部４４に加え、調整部４５を設けている。そしてこの調整部４５において、比較部４４によりデータベース５１から選択された音声フレームデータを、その音声パラメータ値が上記分析部４２より抽出されたマスキング対象音声の音声パラメータ値と一致するように、または近付けるべく調整し、この調整後の音声データをマスキング音としてスピーカ７から拡声出力するようにしている。

【００６９】

したがって、マスキング対象音声と音声パラメータ値が所定の差の範囲内で一致するエントリデータをデータベース５１から発見できなかった場合でも、当該エントリデータの音声データが、マスキング対象音声と音声パラメータ値ができる限り近づくように調整されるため、マスキング対象の音声をさらに効果的にマスキングすることが可能となる。また、このようにマスキング音を生成する第２の制御ユニット４に調整部４５を設けたことにより、先に述べた変換部３３における段階数を減らしてエントリデータ数を削減することができ、これによりデータベース５１の記憶容量を削減すると共に、データベース５１からエントリデータを選択する際のアクセス時間を短縮して、マスキング音の出力遅延を減少させることが可能となる。

【００７０】

さらに本実施形態では、マスキング音を生成する第２の制御ユニット４に予測部４３を備え、この予測部４３において、マスキング用分析部４２により算出された現フレームの音声パラメータのフレーム内平均値と、特徴量バッファ部５３に記憶されている過去のフレームの音声パラメータのフレーム内平均値とをもとに、数フレーム先の音声パラメータのフレーム内平均値を予測する。そして、上記マスキング用分析部４２より算出された現フレームの音声パラメータのフレーム内平均値の代わりに、上記予測された数フレーム先の音声パラメータのフレーム内平均値を比較部４４に供給することも可能にしている。

【００７１】

したがって、例えばマスキング音の遅延量をさらに少なくすることが要求される場合には、上記予測された数フレーム先の音声パラメータのフレーム内平均値を選択することで

、マスキング音の遅延量を減少させて、マスキング効果をさらに向上させることが可能となる。

【００７２】

図７は、Wholeデータベースを用いて現フレームの音声パラメータＰinput のフレーム内平均値をそのまま使用してマスキングを行った場合と、上記予測部４３により予測された音声パラメータＰ＾input のフレーム内平均値を用いてマスキングを行った場合とで、ターゲット音とマスキング音との比（ＴＭＲ）に対する単語了解度を計測した結果の一例を示したものである。

【００７３】

同図から明らかなように、上記二つのマスキング音を比較した場合、それぞれのＴＭＲにおける単語了解度に約２０％の差が見られた。また、単語了解度が４０％となるマスキング音の呈示レベルを比較した（単語了解度４０％という値は、サウンドマスキングシステムのマスキング音を評価する際に頻繁に使用される）ところ、約3dBの差が見られた。この数値は，遅延を想定したマスキング音が遅延を伴わない理想的なマスキング音と同等の性能（同等の単語了解度）を持つために、マスキング音に約１．４倍の音量が必要なことを意味する。以上のことから、予測部４３を使用することで、マスキング音作成処理に伴う遅延によるサウンドマスキングシステムの性能悪化を緩和することができる。

【００７４】

［第２の実施形態］

この発明の第２の実施形態は、データベースに男性話者音声データベースと女性話者音声データベースを設け、マスキング対象の音声に適合するマスキング用音データを上記データベースから読み出す際に、マスキング対象の音データから抽出した基本周波数に応じて上記各データベースを切り替えるようにしたものである。

【００７５】

図８はこの発明の第２の実施形態に係るサウンドマスキング装置の要部構成を示すブロック図である。なお、同図において図１と同一部分には同一符号を付して詳しい説明は省略する。

【００７６】

図８に示すようにデータベース５１０には、男性話者音声データベース５１１と、女性話者音声データベース５１２が設けられている。男性話者音声データベース５１１には、基本周波数が平均的な男性話者の基本周波数範囲に含まれる標本音声に基づいて、第１の制御ユニット３により生成されたマスキング用の音データ群と、当該音データに対応するパラメータ値が、エントリデータとして記憶される。

【００７７】

女性話者音声データベース５１２には、同様に、基本周波数が平均的な女性話者の基本周波数範囲に含まれる標本音声に基づいて、第１の制御ユニット３により生成されたマスキング用の音データ群と、当該音データに対応するパラメータ値が、エントリデータとして記憶される。

【００７８】

なお、標本音声の音声パラメータの分析処理、マスキング用の音データ群の生成処理、及び変換部３３０による変換処理の各手順と内容については、第１の実施形態で述べたWholeデータベースの作成方法が適用される。なお、男性話者音声データベース５１１と、女性話者音声データベース５１２は、別々のデータベースにする必要はなく、１個のデータベースとして構成するようにしてもよい。また反対に、男女それぞれ複数のデータベースを用意してもよい。

【００７９】

一方、第２の制御ユニット４の比較部４４０は、マスキング用分析部４２又は予測部４３から与えられたパラメータのうち、マスキング対象の音データの基本周波数を予め設定した閾値と比較することにより、上記マスキング対象の音データが男性話者のものか或いは女性話者のものかを判定する。

【００８０】
そして、この判定の結果、上記マスキング対象の音データが男性話者であれば、上記男性話者音声データベース５１１を選択し、当該男性話者音声データベース５１１からエントリデータを順次読み出す。そして、パラメータが基本周波数であれば上記マスキング対象音のパラメータ値と最も近いものを選択する。また、パラメータがフォルマントであれば上記マスキング対象音のパラメータ値に対し最も遠いものを選択する。そして、この選択したパラメータ値を調整部４５０に通知する。
【００８１】
調整部４５０は、上記通知されたパラメータ値に関連付けられたマスキング用の音声フレームデータを上記男性話者音声データベース５１１から読み出し、この読み出された音声フレームデータを、その音声パラメータのフレーム内平均値が上記現フレームの音声パラメータのフレーム内平均値、または上記予測された音声パラメータのフレーム内平均値と一致するように調整し、音声出力処理部４６へ出力する。
【００８２】
これに対し、マスキング対象の音データが女性話者と判定されたとする。この場合、上記女性話者音声データベース５１２を選択し、当該女性話者音声データベース５１２からエントリデータを順次読み出す。そして、先に述べた男性話者の場合と同様に、パラメータが基本周波数であれば上記マスキング対象音のパラメータ値と最も近いものを選択する。また、パラメータがフォルマントであれば上記マスキング対象音のパラメータ値に対し最も遠いものを選択する。そして、この選択したパラメータ値を調整部４５０に通知する。
【００８３】
調整部４５０は、上記通知されたパラメータ値に関連付けられたマスキング用の音声フレームデータを上記女性話者音声データベース５１２から読み出し、この読み出された音声フレームデータを、その音声パラメータのフレーム内平均値が上記現フレームの音声パラメータのフレーム内平均値、または上記予測された音声パラメータのフレーム内平均値と一致するように調整し、音声出力処理部４６へ出力する。
【００８４】
このような構成であるから、マスキング対象話者が男性であっても、また女性であっても、それぞれの発話音声の音響特性によりマッチしたマスキング用音データを選択し、マスキングを行うことができる。
【００８５】
［第３の実施形態］
この発明の第３の実施形態は、データベースとして第１の制御ユニット３によりMiddleデータベースを作成し、この作成されたMiddleデータベースを用いてマスキング対象の音データに対しパラメータ値が最適なものを選択し、この選択されたパラメータに対応するマスキング用の音データを出力するようにしたものである。
【００８６】
以下にMiddleデータベースの作成処理手順と処理内容を説明する。なお、この実施形態においても図１に示した構成を用いて説明を行う。
先ず音声入力処理部３１は、標本音声として日本語の単音節音声（清音、濁音、半濁音、拗音）を複数個（例えば１００種類）選択し、それぞれの単音節音声からそのフォルマント遷移部を中心に子音部の末尾部分と母音部の冒頭部分とを含む１フレーム長の音データを抽出する。この抽出された音声フレームを親エントリと呼ぶ。すなわち、親エントリは１音節につき１個生成される。続いてエントリ用分析部３２が、上記単音節音声毎に、上記抽出された１フレーム長の音データの音響特性を分析して当該音響特性を表すパラメータ値、例えば基本周波数の平均値を算出する。
【００８７】
次に、変換部３３が、上記親エントリのそれぞれに対し、所定の変換処理を行って新たな複数の音データを生成し、この新たに生成した複数の音データを子エントリとする。例

えば、各親エントリのそれぞれについてそのフレームの平均基本周波数を、－５０Hz、－４８Hz、…、－２Hz、＋２Hz、＋４Hz、…、＋１００Hzのように変換することにより、複数の子エントリを生成する。そして、上記音節毎に上記親エントリと上記生成された複数の子エントリをデータベース５１に記憶させる。なお、上記子エントリを生成するための変換処理は、基本周波数のピッチ変換に限らず、スペクトル変換等を用いてもよい。

【００８８】

一方、上記Middleデータベースを用いたマスキング用音データの選択処理は以下のように行われる。なお、この実施形態においても図１に示した構成を用いて説明を行う。

すなわち、先ず入力されたマスキング対象の音データ（ターゲット）をリアルタイムに先頭から１００msec長ずつ音声入力処理部４１に取り込み、これによりターゲットの入力音データを複数のフレームに分割する。次にエントリ用分析部４２により、上記分割された各フレームに対して平均基本周波数を計算し、さらにFFT（First Fourier Transform）ケプストラムの低ケフレンシ部（１次～３０次の項）も合わせて計算する。

【００８９】

続いて比較部４４により、上記計算されたターゲットの各フレームにおける平均基本周波数と、データベース５１に記憶された全てのエントリにおける平均基本周波数との差を計算し、ターゲット内の注目するフレームにおける平均基本周波数が近接しているエントリ、例えば差が許容範囲±１Hz以内のものを候補としてすべて選択する。そして、この選択された候補の中から、ターゲットの当該フレームとのスペクトル距離（実際はケプストラム距離、つまりFFTケプストラムの低ケフレンシ部における各次元の差の和）が最も大きいエントリを、そのフレームに対するマスキング用音データとして選択する。

【００９０】

なお、このとき第２の実施形態で述べたように、男性話者音声データベース５１１と女性話者音声データベース５１２が別々に設けられている場合には、マスキング用音データを、男性話者音声のターゲットに対しては男性話者音声データベース５１１から、女性話者音声のターゲットに対しては女性話者音声データベース５１２からそれぞれ選択する。以後、フレーム毎に上記処理を繰り返す。

【００９１】

次に、調整部４５により、上記処理を繰り返すことにより選択された各エントリを順次連結して信号Ａを生成する。なお、上記選択されたエントリを順次連結する際に、ターゲットのレベルにマスキング音のレベルを追従させる。実際には、ターゲットの各フレームと対応するエントリの実効値が等しくなるようにレベルを調節する。

【００９２】

また、上記信号Ａとは別に、上記ターゲットを１／２フレーム遅延させた時点から上記一連の処理を繰り返し行い、これにより信号Ｂを生成する。そして、この作成された信号Ｂと上記作成された信号Ａとを加算し、この加算された信号Ａ＋Ｂをマスキング用の音データとする。このように信号Ａに、位相を１／２フレーム遅延させた信号Ｂを足し合わせたことにより、マスキング音のレベルが下がる区間を減少させることができる。

【００９３】

図９は、上記Middleデータベースに記憶されたマスキング用音データを用いて単語了解度試験を行った結果を、Wholeデータベースに記憶されたマスキング用音データを用いて同様の試験を行った結果と対比して示したものである。同図から明らかなように、WholeとMiddleとの間には性能の差が見られなかった。

【００９４】

以上述べたように第３の実施形態によれば、単音節ごとにフォルマント遷移部を中心に子音部の末尾部分と母音部の冒頭部分とを含む領域のみについて、マスキング用音データのエントリデータ群を生成してMiddleデータベースを作成したことにより、単音節毎にその全領域を複数のフレームに分割し、これらのフレーム毎にマスキング用音データのエントリデータ群を生成してデータベースにエントリする場合に比べ、データベースへのエントリデータ量とエントリに必要な処理時間を大幅に減らすことができ、さらにデータベー

スからマスキング対象の音データに適したマスキング用音データを選択するために必要な時間を短縮して、マスキング処理の応答性を高めることができる。

【００９５】

［その他の実施形態］

前記実施形態では、データベースにエントリデータを登録する際に、変換部において任意の一人の音声をもとにその音声パラメータ値の異なる複数のマスキング用の音データを作成し登録するようにした。しかし、それに限らず複数の人の音声をもとにそれぞれパラメータ値の異なる複数のマスキング用の音データを作成し登録するようにしてもよく、それに加えて環境音や定常雑音等をもとにパラメータ値の異なる複数のマスキング用の音データを作成し登録するようにしてもよい。

【００９６】

また、前記実施形態では音声パラメータのフレーム内平均値を算出し、このフレーム内平均値を段階的に変化させてマスキング用の音データを作成したが、フレーム内平均値に限定されることなく、フレーム内のピーク値や中央値を段階的に変化させてマスキング用の音データを作成するようにしてもよい。

さらに、マスキング用の音データを作成する際に、時間反転処理を含むその他の処理を施したマスキング用の音データを作成するようにしてもよい。

【００９７】

また、第１の制御ユニット、第２の制御ユニット及び記憶ユニットを１つの装置内に設けずに、それぞれ別の装置として独立して設けてもよい。また、第１の制御ユニットと記憶ユニットとを１つの装置とし第２の制御ユニットを別の装置として設けたり、第２の制御ユニットと記憶ユニットとを１つの装置とし第１の制御ユニットを別の装置として設けてもよい。何れも場合も、各装置間の接続は、通信回線や信号ケーブルを介して行われる。

【００９８】

さらに、記憶ユニットについてはクラウドコンピュータ上に設けるようにし、別々の場所に設けられた複数の第１及び第２の制御ユニットがインターネット等のネットワークを介して上記記憶ユニットにアクセスするようにしてもよい。このようにすると１台の記憶ユニットを複数の第１及び第２の制御ユニットにより共有することができる。

【００９９】

その他、サウンドマスキング装置の構成や、エントリデータの作成処理、オンラインにおけるマスキング音の選択・生成処理の手順及び処理内容、入力音声のフレーム長等についても、この発明の要旨を逸脱しない範囲で種々変形して実施可能である。

【０１００】

要するにこの発明は、上記実施形態そのままに限定されるものではなく、実施段階ではその要旨を逸脱しない範囲で構成要素を変形して具体化できる。また、上記実施形態に開示されている複数の構成要素の適宜な組み合せにより種々の発明を形成できる。例えば、実施形態に示される全構成要素から幾つかの構成要素を削除してもよい。さらに、異なる実施形態に亘る構成要素を適宜組み合せてもよい。

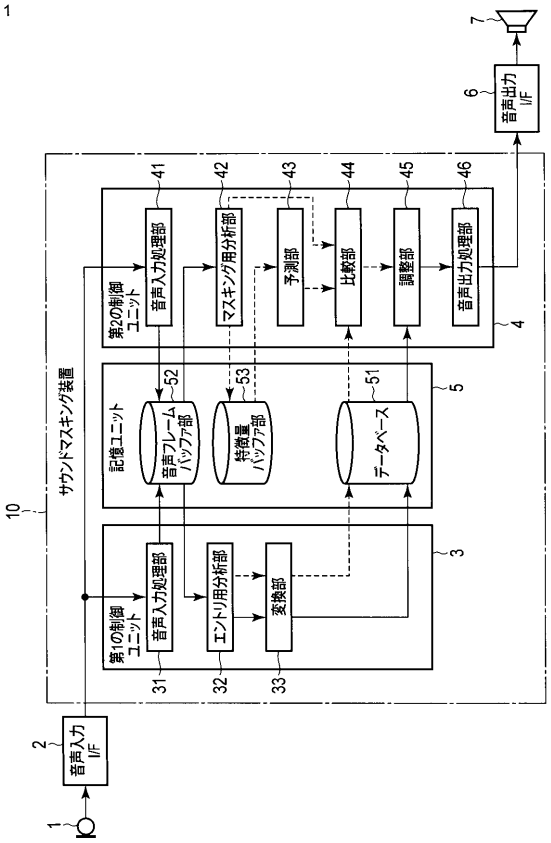
【符号の説明】

【０１０１】

１…マイクロホン、２…音声入力インタフェース（音声入力Ｉ／Ｆ）、３…第１の制御ユニット、４…第２の制御ユニット、５…記憶ユニット、６…音声出力インタフェース（音声出力Ｉ／Ｆ）、７…スピーカ、１０…サウンドマスキング装置、３１…音声入力処理部、３２…エントリ用分析部、３３，３３０…変換部、４１…音声入力処理部、４２…マスキング用分析部、４３…予測部、４４，４４０…比較部、４５，４５０…調整部、４６…音声出力処理部、５１，５１０…データベース、５２…音声フレームバッファ部、５３…特徴量バッファ部、５１１…男性話者音声データベース、５１２…女性話者音声データベース。

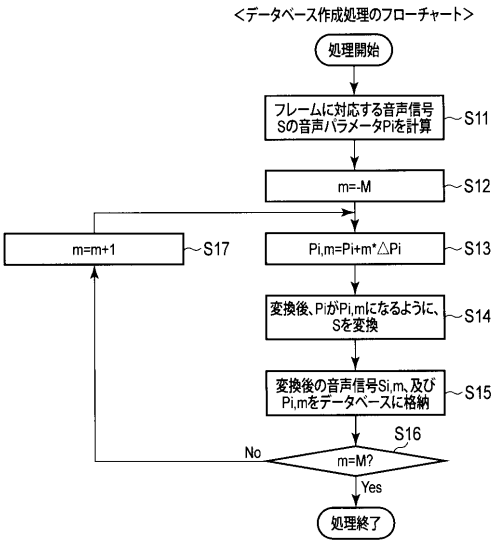
【図１】

図1



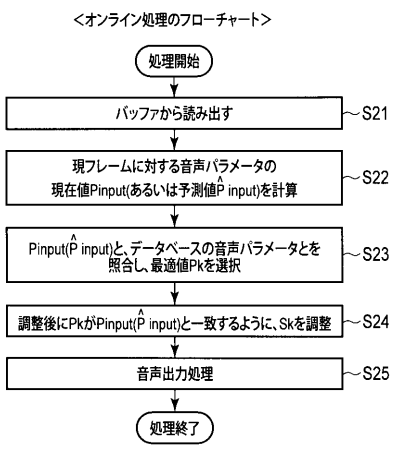
【図２】

図2



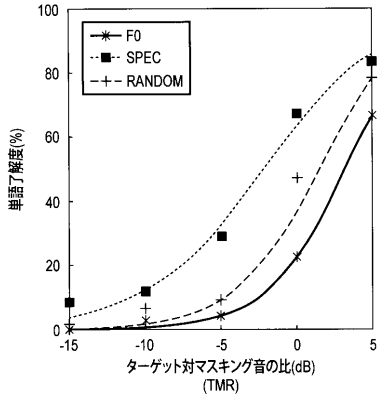
【図３】

図3



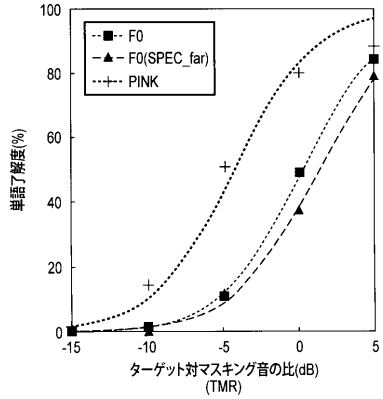
【図４】

図4



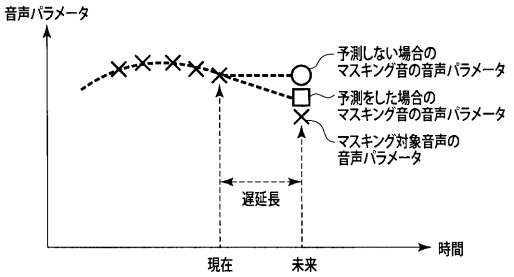
【図５】

図5



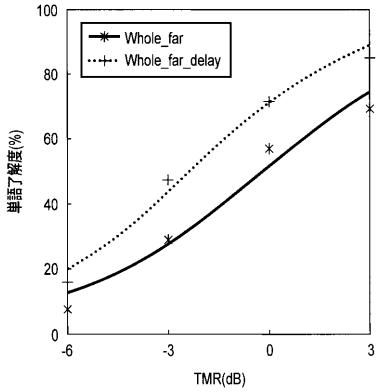
【図６】

図６



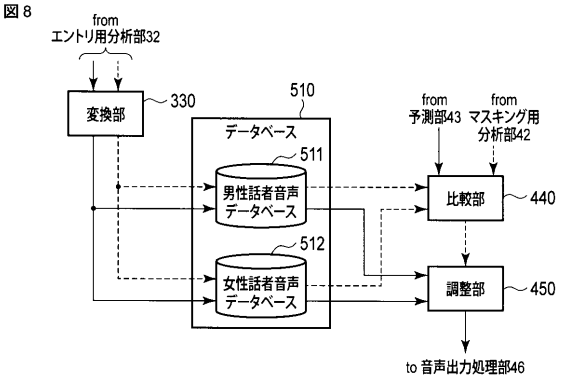
【図７】

図７



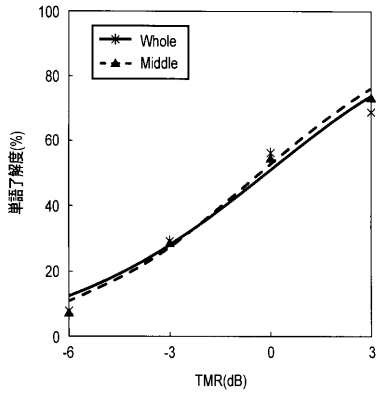
【図８】

図８



【図９】

図９



フロントページの続き

(74)代理人 100158805
弁理士 井関 守三
(74)代理人 100179062
弁理士 井上 正
(74)代理人 100124394
弁理士 佐藤 立志
(74)代理人 100112807
弁理士 岡田 貴志
(74)代理人 100111073
弁理士 堀内 美保子
(72)発明者 荒井 隆行
東京都千代田区紀尾井町７番１号 学校法人 上智学院 上智大学 理工学部 情報理工学科内
(72)発明者 三戸 武大
東京都千代田区紀尾井町７番１号 学校法人 上智学院 上智大学大学院 理工学研究科内
(72)発明者 安 啓一
東京都千代田区紀尾井町７番１号 学校法人 上智学院 上智大学 理工学部 情報理工学科内
Ｆターム(参考) 5D061 FF10