

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第6349112号
(P6349112)

(45) 発行日 平成30年6月27日 (2018. 6. 27)

(24) 登録日 平成30年6月8日 (2018. 6. 8)

(51) Int. Cl.	F I
G 1 O K 11/175 (2006. 01)	G 1 O K 11/175
G 1 O L 19/00 (2013. 01)	G 1 O L 19/00 3 0 0
G 1 O L 21/013 (2013. 01)	G 1 O L 21/013

請求項の数 13 (全 22 頁)

(21) 出願番号	特願2014-48187 (P2014-48187)	(73) 特許権者	502350504
(22) 出願日	平成26年3月11日 (2014. 3. 11)		学校法人上智学院
(65) 公開番号	特開2014-199445 (P2014-199445A)		東京都千代田区紀尾井町7番1号
(43) 公開日	平成26年10月23日 (2014. 10. 23)	(74) 代理人	100108855
審査請求日	平成29年3月2日 (2017. 3. 2)		弁理士 蔵田 昌俊
(31) 優先権主張番号	特願2013-48473 (P2013-48473)	(74) 代理人	100109830
(32) 優先日	平成25年3月11日 (2013. 3. 11)		弁理士 福原 淑弘
(33) 優先権主張国	日本国 (JP)	(74) 代理人	100103034
			弁理士 野河 信久
特許法第30条第2項適用 2012年秋季研究発表会		(74) 代理人	100075672
講演論文集講演要旨・講演論文のCD-ROM (平成			弁理士 峰 隆司
24年9月11日) 一般社団法人日本音響学会発行第9		(74) 代理人	100153051
1頁、第1243-1246頁に発表			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) 【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、当該音データに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

10

20

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

10

20

30

40

50

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的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。

10

【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、

20

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

30

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

40

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

50

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

10

20

30

40

50

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ーを保護する手法の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ーを保護するために、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、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に変化した場合でも、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に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おいて、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。

【 0 0 0 8 】

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。

【 0 0 0 9 】

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。

【 0 0 1 0 】

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。

【 0 0 1 1 】

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。

【 0 0 1 2 】

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】

【 0 0 1 3 】

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の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。

10

【 0 0 1 4 】

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。

20

【 0 0 1 5 】

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。

【 0 0 1 6 】

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。

30

【 0 0 1 7 】

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。

40

【 0 0 1 8 】

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。

【 0 0 1 9 】

50

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に変化した場合でも、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に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】

10

【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。

20

【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】

30

【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。

40

【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

50

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 5 2 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 5 3 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 4 3 による特徴量予測処理のために保存する。

【 0 0 2 5 】

第 1 及び第 2 の制御ユニット 3 , 4 はいずれも C P U (Central Processing Unit) 及び D S P (Digital Signal Processor) を備える。なお、これらの C P U 及び D S P 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。

10

【 0 0 2 6 】

第 1 の制御ユニット 3 は、データベース 5 1 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 3 1 と、エントリ用分析部 3 2 と、変換部 3 3 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。

【 0 0 2 7 】

音声入力処理部 3 1 は、上記音声入力 I / F 2 から標本用のデジタル音声信号を受け取り、この受け取ったデ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 5 2 に保存させる。1 フレーム長は例えば 1 0 0 ms に設定されるが、その他の長さに設定してもよい。

20

【 0 0 2 8 】

エントリ用分析部 3 2 は、上記音声フレームバッファ部 5 2 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 F_0 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 3 2 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。

【 0 0 2 9 】

変換部 3 3 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 5 1 に格納する処理を行う。

30

【 0 0 3 0 】

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 4 1 と、マスキング用分析部 4 2 と、予測部 4 3 と、比較部 4 4 と、調整部 4 5 と、音声出力処理部 4 6 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。

【 0 0 3 1 】

音声入力処理部 4 1 は、上記音声入力 I / F 2 からマスキング対象のデジタル音声信号を受け取り、この受け取ったデジタル入力音声信号を上記標本用のデ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 5 2 に保存させる。

40

【 0 0 3 2 】

マスキング用分析部 4 2 は、上記音声フレームバッファ部 5 2 からマスキング対象のデ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 3 2 と同様に、基本周波数 F_0 と、スペクトル特性と、強度からなる。エントリ用分析部 3 2 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 5 3 に一時保存させる処理を行う。

50

【 0 0 3 3 】

予測部 4 3 は、上記マスキング用分析部 4 2 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 5 3 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。

【 0 0 3 4 】

比較部 4 4 は、上記マスキング用分析部 4 2 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 4 3 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 5 1 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 5 1 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。

10

【 0 0 3 5 】

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。

20

【 0 0 3 6 】

調整部 4 5 は、上記比較部 4 4 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 4 2 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。

【 0 0 3 7 】

音声出力処理部 4 6 は、上記調整部 4 5 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 / F 6 へ出力する処理を行う。

30

【 0 0 3 8 】

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 C P U 又は D S P に実行させることにより実現される。

【 0 0 3 9 】

(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 1 0 の動作を説明する。

40

(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 / F 2 でデジタル信号に変換されたのちサウンドマスキング装置 1 0 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。

【 0 0 4 0 】

サウンドマスキング装置 1 0 では、上記入力されたデ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 3 1 に所定のフレーム長 (例えば 1 0 0 ms) 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 5 2 に一時保存される。すなわち、この処理により入力デジタル音声信号は 1 0 0 ms のフレーム長に分割

50

される。

【 0 0 4 1 】

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。

【 0 0 4 2 】

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 P_i 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 P_i として、例えば基本周波数 F_0 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 P_i （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。

【 0 0 4 3 】

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 P_i 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（ $m = -M \sim M$ ））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 $P_{i,m}$ が

$$P_{i,m} = P_i + m \times \Delta P_i$$

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

【 0 0 4 4 】

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 P_i が上記計算された段階 $-M$ における音声パラメータ $P_{i,m}$ 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 $S_{i,m}$ と上記音声パラメータ $P_{i,m}$ 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。

【 0 0 4 5 】

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。

【 0 0 4 6 】

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 F_0 の平均値に対し $\pm 2.5\text{ Hz}$ 、 $\pm 5.0\text{ Hz}$ 、 $\pm 7.5\text{ Hz}$ 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 F_0 が上記 -7.5 Hz 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 $F_0 - 7.5\text{ Hz}$ 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 F_0 が上記 -5.0 Hz のときの音声フレームデータに変換され、 $F_0 - 5.0\text{ Hz}$ 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 F_0 が上記 -2.5 Hz 、 $+2.5\text{ Hz}$ 、 $+5.0\text{ Hz}$ 、 $+7.5\text{ Hz}$ 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。

【 0 0 4 7 】

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。

【 0 0 4 8 】

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 P_i 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 $S_{i,m}$ がそれぞれエントリデータとして登録される。

【 0 0 4 9 】

例えば、音声パラメータ P_i が先に述べた 3 種類 ($i=1,2,3$) の場合であれば、 m ($-M \sim M$) の各段階ごとに、基本周波数 F_0 、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 F_0 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 F_0 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 F_0 、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、 $7 \times 6 = 42$ 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。

【 0 0 5 0 】

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。

【 0 0 5 1 】

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 100 ms）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。

【 0 0 5 2 】

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 -50 Hz , -48 Hz , ... , -2 Hz , $+2\text{ Hz}$, $+4\text{ Hz}$, ... , $+100\text{ Hz}$ ）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。

【 0 0 5 3 】

（ 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 / F 2 でデ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。

【 0 0 5 4 】

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 100 ms）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

10

20

30

40

50

音声フレームバッファ部 5 2 に一時保存される。

【 0 0 5 5 】

次に第 2 の制御ユニット 4 では、マスキング用分析部 4 2、予測部 4 3、比較部 4 4、調整部 4 5 及び音声出力処理部 4 6 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。

【 0 0 5 6 】

すなわち、先ずステップ S 2 1 において、上記音声フレームバッファ部 5 2 から入力音声データ S_k が 1 フレームずつ読み込まれる。そしてステップ S 2 2 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 P_{input} 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 3 2 と同様に、音声パラメータ P_{input} として、基本周波数 F_0 、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 P_i (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 P_{input} としては、基本周波数 F_0 、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。

【 0 0 5 7 】

また、ステップ S 2 2 において予測部 4 3 では、上記マスキング用分析部 4 2 から上記算出された現フレームの音声パラメータ P_{input} 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 P_{input} のフレーム内平均値と、上記特徴量バッファ部 5 3 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ともに、数フレーム先の音声パラメータ $P^{\wedge}_{input}$ のフレーム内平均値が予測される。

【 0 0 5 8 】

次にステップ S 2 3 において、比較部 4 4 の制御の下、上記マスキング用分析部 4 2 で算出された現フレームの音声パラメータ P_{input} のフレーム内平均値、または上記予測部 4 3 により予測された音声パラメータ $P^{\wedge}_{input}$ のフレーム内平均値が、データベース 5 1 に格納されている各エントリデータの音声パラメータ $P_{i,m}$ のフレーム内平均値と順次比較される。

【 0 0 5 9 】

そして、音声パラメータが例えば基本周波数 F_0 の場合或いは音量レベルの場合には、データベース 5 1 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 P_{input} 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 $P^{\wedge}_{input}$ のフレーム内平均値と最も類似する音声パラメータ P_k 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 5 1 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 P_{input} 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 $P^{\wedge}_{input}$ 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 P_k のフレーム内平均値が選択される。

【 0 0 6 0 】

ところで、上記予測部 4 3 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 4 3 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 4 2 では、一定間隔 (例えば 20 ms) 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 5 3 に格納される。予測部 4 3 では、特徴量バッファ部 5 3 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 4 1 から音声出力処理部 4 6 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。

【 0 0 6 1 】

比較部 4 4 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 4 4 では、予測

10

20

30

40

50

部 4 3 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。

【 0 0 6 2 】

続いてステップ S 2 4 において、調整部 4 5 の制御の下で、上記選択された音声パラメータ P_k のフレーム内平均値に対応する音声フレームデータ S_k がデータベース 5 1 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 S_k が、その音声パラメータ P_k 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 P_{input} 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。

【 0 0 6 3 】

最後にステップ S 2 5 において、音声出力処理部 4 6 の制御の下、上記調整部 4 5 により調整された音声フレームデータ S_k が時系列に従い接続されて連続するデジタル音声信号が生成され、音声出力 I / F 6 へ出力される。このデジタル音声信号は、音声出力 I / F 6 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ーは保護される。

【 0 0 6 4 】

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）に対する単語理解度（%）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 S P E C 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 5 1 内のエントリデータを無作為に選択した R A N D O M 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理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。

【 0 0 6 5 】

また図 5 には、TMR と単語理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 _ S P E C 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 P i n k とを比較すると、P i n k 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 _ S P E C マスキング音の方が単語理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。

【 0 0 6 6 】

（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 5 1 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 3 1 及び分析部 3 2 に加え変換部 3 3 を設けている。そしてこの変換部 3 3 において、分析部 3 2 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 5 1 に格納するようにしている。

【 0 0 6 7 】

したがって、データベース 5 1 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の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

10

20

30

40

50

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ともに、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

10

20

30

40

50

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。

10

【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。

20

【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。

30

【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。

40

【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

50

。

【 0 0 8 3 】

調整部 4 5 0 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 5 1 2 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 4 6 へ出力する。

【 0 0 8 4 】

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。

【 0 0 8 5 】

〔 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。

【 0 0 8 6 】

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 3 1 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 1 0 0 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 3 2 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。

【 0 0 8 7 】

次に、変換部 3 3 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、 -50 Hz 、 -48 Hz 、 $\dots$ 、 -2 Hz 、 $+2\text{ Hz}$ 、 $+4\text{ Hz}$ 、 $\dots$ 、 $+100\text{ Hz}$ 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 5 1 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。

【 0 0 8 8 】

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 1 0 0 msec 長ずつ音声入力処理部 4 1 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 4 2 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～3 0 次の項）も合わせて計算する。

【 0 0 8 9 】

続いて比較部 4 4 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 5 1 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 $\pm 1\text{ Hz}$ 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

10

20

30

40

50

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。

【 0 0 9 0 】

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 5 1 1 と女性話者音声データベース 5 1 2 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 5 1 1 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 5 1 2 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。

【 0 0 9 1 】

次に、調整部 4 5 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。

【 0 0 9 2 】

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。

【 0 0 9 3 】

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理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。

【 0 0 9 4 】

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。

【 0 0 9 5 】

〔 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。

【 0 0 9 6 】

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。

【 0 0 9 7 】

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

10

20

30

40

50

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わ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わせてもよい。

【符号の説明】

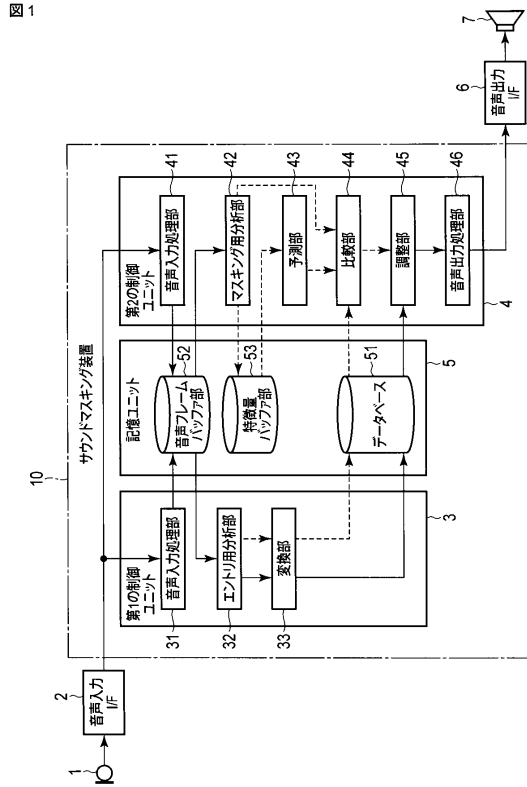
【0101】

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。

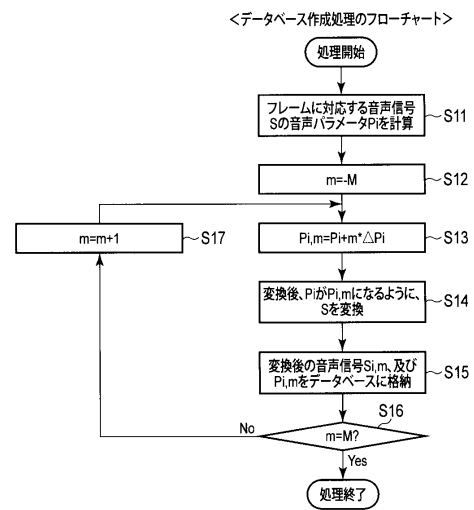
10

20

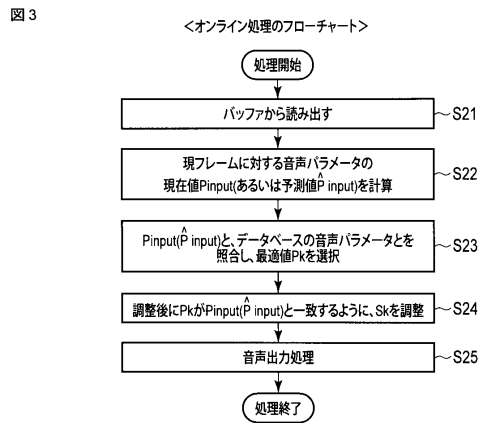
【図 1】



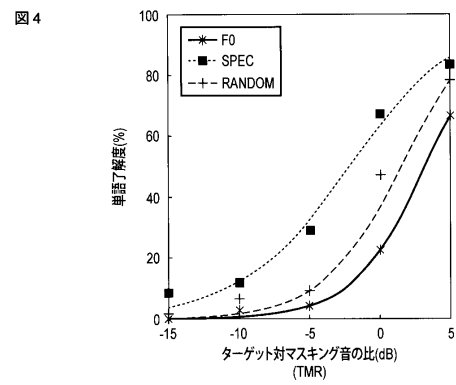
【図 2】



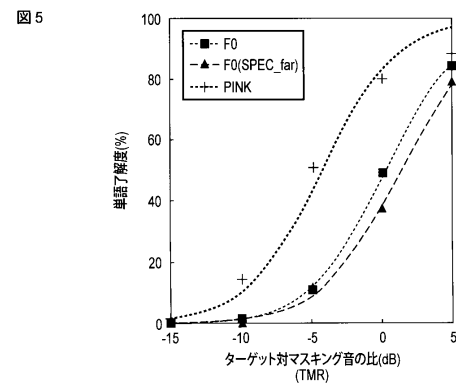
【図 3】



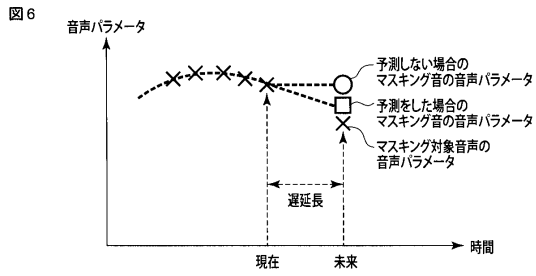
【図 4】



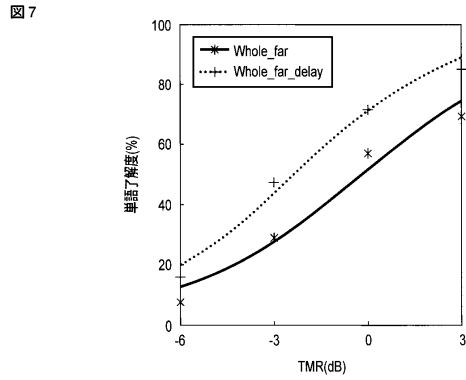
【図 5】



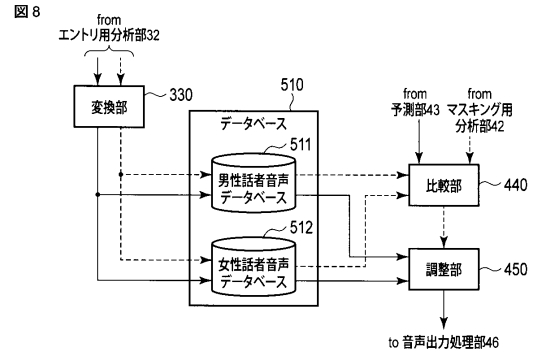
【図 6】



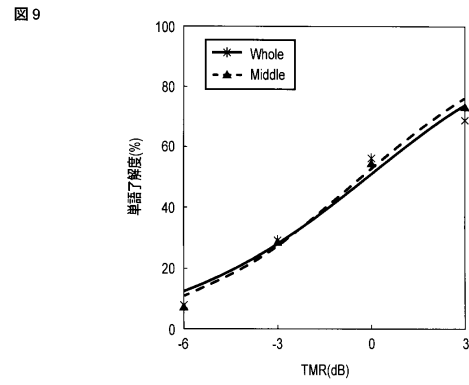
【図 7】



【図 8】



【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

審査官 上田 雄

(56)参考文献 特開2008-233672(JP,A)

特開2001-83979(JP,A)

(58)調査した分野(Int.Cl., DB名)

G10K 11/00 - 11/36

G10L 13/00 - 25/93