

(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2016-186516

(P2016-186516A)

(43) 公開日 平成28年10月27日(2016.10.27)

(51) Int.Cl.	F I	テーマコード (参考)
G 1 O L 15/065 (2013.01)	G 1 O L 15/065 Z	
G 1 O L 13/033 (2013.01)	G 1 O L 13/033 1 O 2 Z	
G 1 O L 13/08 (2013.01)	G 1 O L 13/08 1 5 O B	
G 1 O L 15/06 (2013.01)	G 1 O L 15/06 3 O O Y	

審査請求 未請求 請求項の数 6 O L (全 10 頁)

(21) 出願番号 特願2015-65788 (P2015-65788)  
 (22) 出願日 平成27年3月27日 (2015.3.27)

(71) 出願人 000004226  
 日本電信電話株式会社  
 東京都千代田区大手町一丁目5番1号  
 (74) 代理人 100121706  
 弁理士 中尾 直樹  
 (74) 代理人 100128705  
 弁理士 中村 幸雄  
 (74) 代理人 100147773  
 弁理士 義村 宗洋  
 (72) 発明者 芦原 孝典  
 東京都千代田区大手町一丁目5番1号 日  
 本電信電話株式会社内  
 (72) 発明者 浅見 太一  
 東京都千代田区大手町一丁目5番1号 日  
 本電信電話株式会社内

最終頁に続く

(54) 【発明の名称】 疑似音声信号生成装置、音響モデル適応装置、疑似音声信号生成方法、およびプログラム

(57) 【要約】

【課題】 学習データが十分に入手できない場合であっても、認識率を向上させる。

【解決手段】 疑似音声信号生成部 1 2 は、出現頻度が低い音声現象の音響特性を表現した変換ルールに基づいて入力音声信号の音源を変換し、その音源で入力音声信号から抽出したスペクトル包絡を励振することで疑似音声信号を生成する。疑似音響特徴量生成部 1 3 は、疑似音声信号から音響特徴量を抽出し疑似音響特徴量系列を生成する。音響モデル学習部 1 4 は、疑似音響特徴量系列を用いて音響モデルを学習する。

【選択図】 図 1

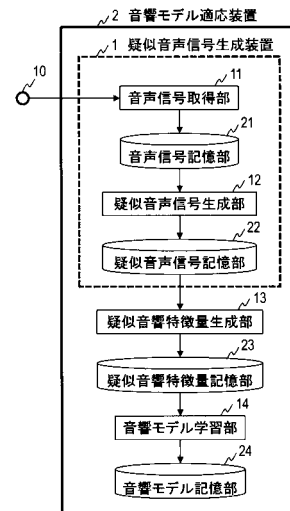


図 1

## 【特許請求の範囲】

## 【請求項 1】

出現頻度が低い音声現象の音響特性を表現した変換ルールに基づいて入力音声信号の音源を変換し、当該音源で上記入力音声信号から抽出したスペクトル包絡を励振することで疑似音声信号を生成する疑似音声信号生成部を含む疑似音声信号生成装置。

## 【請求項 2】

請求項 1 に記載の疑似音声信号生成装置であって、

上記疑似音声信号生成部は、上記入力音声信号の音源をホワイトノイズに変換し、当該音源で上記スペクトル包絡を励振することによって上記疑似音声信号を生成するものである

疑似音声信号生成装置。

## 【請求項 3】

請求項 1 に記載の疑似音声信号生成装置であって、

上記疑似音声信号生成部は、上記入力音声信号の音源を母音区間の基本周波数が低くなり断続的に音声が消滅するように変換し、当該音源で上記入力音声信号から抽出したスペクトル包絡を励振することによって上記疑似音声信号を生成するものである

疑似音声信号生成装置。

## 【請求項 4】

請求項 1 から 3 のいずれかに記載の疑似音声信号生成装置により生成した疑似音声信号を記憶する疑似音声信号記憶部と、

上記疑似音声信号から音響特徴量を抽出し疑似音響特徴量系列を生成する疑似音響特徴量生成部と、

上記疑似音響特徴量系列を用いて音響モデルを学習する音響モデル学習部と、

を含む音響モデル適応装置。

## 【請求項 5】

疑似音声信号生成部が、出現頻度が低い音声現象の音響特性を表現した変換ルールに基づいて入力音声信号の音源を変換し、当該音源で上記入力音声信号から抽出したスペクトル包絡を励振することで疑似音声信号を生成する疑似音声信号生成ステップを含む疑似音声信号生成方法。

## 【請求項 6】

請求項 1 から 3 のいずれかに記載の疑似音声信号生成装置もしくは請求項 4 に記載の音響モデル適応装置としてコンピュータを機能させるためのプログラム。

## 【発明の詳細な説明】

## 【技術分野】

## 【0001】

この発明は、音声認識技術に関し、特に、音声認識のタスクに音響モデルを適応させるために用いる学習データを生成する技術に関する。

## 【背景技術】

## 【0002】

特許文献 1 には、音声認識において実用レベルの性能を担保するために、音声認識の対象とするタスク（以下、認識対象タスクと呼ぶ）に対して音響モデルを適応させる技術が記載されている。ここで、認識対象タスクとは、元々の音響モデルに対して、話者や雑音タイプ、喋り方などの音響的特徴が異なるタスクである。

## 【0003】

一般的に、音声認識の性能は認識対象タスクの学習データ量に依存して上下する。つまり、認識対象タスクの学習データが満足に存在しない状況で、従来の技術により音響モデルを適応させたとしても満足のいく認識率は得られない場合が多い。そこで通常は、認識対象タスクの音声を十分に集め、その音声を書き起こしすることで所望の量の学習データを収集するのであるが、そのためには莫大な金銭的・時間的コストを要する。また、認識対象タスクの音声が入手可及であるならば、書き起こしによる学習データの収集を

10

20

30

40

50

実施することが可能だが、そもそもあらゆるタスクにおいて十分な量の音声が入手可能というわけではない。例えば、方言や日本人が英語を話す音声など、十分な量の音声を入手することが難しいタスクも存在する。

【0004】

認識対象タスクのデータベースを所有していたとしても、認識対象タスク内で出現頻度の少ない音声現象が存在する場合、その音声現象に対しても頑健な音響モデルを構築するためには、その音声現象の学習データも十分な量を収集する必要がある。例えば、人間同士の自然な会話では、様々な種類の発声を発話の一部としており、「ささやくような発声（以降、「ささやき声」と呼ぶ）」や「低周波数でのブツブツした音になる発声（以降、「ボーカルフライ」と呼ぶ）」等の現象が存在している。「ささやき声」とは声帯振動を伴わない発声により生成される音声を指す。例えば、周囲に声を漏らさないためにコソコソ話す際（例えば、公の場でモバイル端末に話しかける場合等）に、しばしば現れる発声である。「ボーカルフライ」とは「きしみ声」や「エッジボイス」とも呼ばれ、声帯声門がわずかに開き緩やかな声帯震動により生成される低周波数の音声を指す。例えば、議論の場において頭の中で考えながら発話する場合や少し自信が無くなった場合等に、発話の全体もしくはその一部に出現する発声である。このようなささやき声やボーカルフライは、通常の発声に比べると圧倒的に頻度が少ないため、さまざまな話者で音響モデルの学習をするために十分な量を収集することは困難である。したがって、ささやき声やボーカルフライ等で発声された発話は誤認識となる可能性が高くなってしまう。

10

【0005】

ささやき声やボーカルフライが通常の発話とどのように音響特性が異なるのかについては、非特許文献1や非特許文献2が詳しい。なお、非特許文献1でもささやき声を学習することで認識精度の改善を実現しているが、ここでは既に学習データを十分所有している場合を想定している。

20

【0006】

非特許文献3には、声道長正規化（VTLN: Vocal Tract Length Normalization）のWarping Factorを複数の値で実行することで、学習データにおける話者バリエーションを疑的に作成する方法が記載されている。なお、VTLNについては非特許文献4に記されている。

30

【先行技術文献】

【特許文献】

【0007】

【特許文献1】特開2007-249051号公報

【非特許文献】

【0008】

【非特許文献1】伊藤太介，武田一哉，板倉文忠，“ささやき声の音響分析と音声認識への応用”，信学技報，DSP2001-98，SP2001-71，pp. 59-64，2001

【非特許文献2】M. Blomgren, Y. Chen, M. L. Ng, H. R. Gilbert, “Acoustic, aerodynamic, physiologic, and perceptual properties of modal and vocal fry registers”, Journal of the Acoustical Society of America, vol. 103, pp. 2649-2658, 1998

40

【非特許文献3】N. Jaitly, G. E. Hinton, “Vocal Tract Length Perturbation (VTLN) improves speech recognition”, ICML Workshop on Deep Learning for Audio, Speech, and Language Processing, 2013

【非特許文献4】E. Eide, H. Gish, “A parametric approach to vocal tract length normalization”, ICASSP, pp. 346-348, 1996

【発明の概要】

【発明が解決しようとする課題】

【0009】

しかしながら、非特許文献3に記載の従来技術では、話者の声質を変換し、話者のバリエーションを拡充することのみを目的としており、ささやき声やボーカルフライのような

50

出現頻度の少ない音声現象に関する学習データを疑似生成する場合には利用することができない。

【0010】

この発明の目的は、ささやき声やボーカルフライのような出現頻度が少ない音声現象を認識対象タスクとする音響モデル適応において、十分な量の学習データを収集できない状況であっても、認識率を向上させることである。

【課題を解決するための手段】

【0011】

上記の課題を解決するために、この発明の疑似音声信号生成装置は、出現頻度が低い音声現象の音響特性を表現した変換ルールに基づいて入力音声信号の音源を変換し、その音源で入力音声信号から抽出したスペクトル包絡を励振することで疑似音声信号を生成する疑似音声信号生成部を含む。

10

【発明の効果】

【0012】

この発明の疑似音声生成技術は、ささやき声やボーカルフライのような出現頻度が少ない音声現象を認識対象タスクとする場合に、学習データが十分に入手できない状況下であっても、ルールベースに基づいて認識対象タスクの音声信号を疑似生成し、その疑似音声信号を用いて音響モデルを適応させる。これにより、ささやき声やボーカルフライに頑健な音響モデルを生成でき、認識率を向上することができる。

【図面の簡単な説明】

20

【0013】

【図1】図1は、疑似音声信号生成装置および音響モデル適応装置の機能構成を例示する図である。

【図2】図2は、疑似音声信号生成方法および音響モデル適応方法の処理フローを例示する図である。

【発明を実施するための形態】

【0014】

この発明では、ささやき声やボーカルフライのような出現頻度が少ない音声現象を認識対象タスクとする音響モデル適応において、音響モデルを適応させるために十分な量の学習データを収集できない状況を想定する。まず、ささやき声またはボーカルフライと通常の発声との音響特性の違いを予め変換ルールで定めておく。次に、その変換ルールに基づいて通常の発声による学習データからささやき声またはボーカルフライによる学習データを疑似的に生成する。そして、その疑似学習データを用いて音響モデルを適応させる。

30

【0015】

以下、この発明の実施の形態について詳細に説明する。なお、図面中において同じ機能を有する構成部には同じ番号を付し、重複説明を省略する。

【0016】

[第一実施形態]

第一実施形態では、ルールベースによりささやき声の音声信号を疑似生成し、その疑似音声信号を用いて音響モデルを適応する方法について説明する。変換ルールはささやき声の音響特性に基づいて事前に作成しておき、その変換ルールに基づいて通常の発声からささやき声に変換することで、ささやき声による音声信号の疑似生成を行う。

40

【0017】

第一実施形態の疑似音声信号生成装置1は、図1に例示するように、入力端子10、音声信号取得部11、疑似音声信号生成部12、音声信号記憶部21、および疑似音声信号記憶部22を含む。

【0018】

第一実施形態の音響モデル適応装置2は、図1に例示するように、疑似音声信号生成装置1の各構成部に加えて、疑似音響特徴量生成部13、音響モデル学習部14、疑似音響特徴量記憶部23、および音響モデル記憶部24を含む。図1では、音響モデル適応装置

50

2に疑似音声信号生成装置1のすべての構成部が含まれる構成を例示したが、疑似音声信号生成装置1が生成した疑似音声信号を記憶させた疑似音声信号記憶部22のみを含む構成とすることも可能である。

【0019】

疑似音声信号生成装置1および音響モデル適応装置2の各装置は、例えば、中央演算処理装置(CPU: Central Processing Unit)、主記憶装置(RAM: Random Access Memory)などを有する公知又は専用のコンピュータに特別なプログラムが読み込まれて構成された特別な装置である。各装置は、例えば、中央演算処理装置の制御のもとで各処理を実行する。各装置に入力されたデータや各処理で得られたデータは、例えば、主記憶装置に格納され、主記憶装置に格納されたデータは必要に応じて読み出されて他の処理に利用される。また、各装置が備える各処理部の少なくとも一部が集積回路等のハードウェアによって構成されていてもよい。

10

【0020】

疑似音声信号生成装置1および音響モデル適応装置2が備える各記憶部は、例えば、RAM(Random Access Memory)などの主記憶装置、ハードディスクや光ディスクもしくはフラッシュメモリ(Flash Memory)のような半導体メモリ素子により構成される補助記憶装置、またはリレーショナルデータベースやキーバリューストアなどのミドルウェアにより構成することができる。各装置が備える各記憶部は、それぞれ論理的に分割されていればよく、一つの物理的な記憶装置に記憶されていてもよい。

20

【0021】

図2を参照して、第一実施形態の疑似音声信号生成方法の処理手続きを説明する。

【0022】

ステップS10において、入力端子10へ、学習データとする音声信号が入力される。学習データの音声信号には、認識対象タスクであるささやき声による音声信号と、通常の発声による音声信号とが含まれる。入力される音声信号は、マイクロホン等の收音手段を入力端子10へ接続してリアルタイムに人間の発話を收音したものであってもよいし、あらかじめ人間の発話をICレコーダーやスマートフォンの録音機能のような録音手段で不揮発性メモリやハードディスクドライブのような記録媒体へ録音し、入力端子10へ接続した再生手段により再生することで入力してもよい。

30

【0023】

ステップS11において、音声信号取得部11は、アナログの入力音声信号をデジタル信号に変換する。入力端子10からデジタルの音声信号が入力される場合には、音声信号取得部11は備えなくともよい。デジタルの入力音声信号は、音声信号記憶部21へ記憶される。

【0024】

ステップS12において、疑似音声信号生成部12は、音声信号記憶部21に記憶されたデジタルの入力音声信号を読み込み、予め定められた変換ルールに基づいて入力音声信号の音響特性を変換してささやき声の疑似音声信号を生成する。生成した疑似音声信号は、疑似音声信号記憶部22へ記憶される。

40

【0025】

変換ルールはささやき声と通常の発声との違いを表現するようなルールである。取り得る変換ルールとしては、例えば、音声信号からスペクトル包絡情報を抽出し、音源のみをホワイトノイズ化してスペクトル包絡情報(声道特性フィルタ)を励振させ、有声音を無声化する(駆動音源をホワイトノイズにすることで模擬的な呼気音とする)ことが考えられる。ここで、スペクトル包絡の推定方法としては、公知のヒルベルト変換や、下記参考文献1に記載されるメル一般化ケプストラム分析などを用いることができる。なお、音声認識に用いる音響特徴量(もしくは、後述の疑似音響特徴量生成部13で出力される音響特徴量系列)がメル周波数ケプストラム係数(MFCC: Mel-Frequency Cepstrum Coefficient)であるならば、駆動音源のみが変更される(すなわち、スペクトル包絡は変わらない)場合、理論上音響特徴量の値は変わらない。したがってこのようなケースでは、MFCCで

50

はない音響特徴量を利用する。例えば、メルフィルタバンクから出力される値をそのまま音声認識の音響特徴量に利用する等が考えられる。

〔参考文献1〕徳田恵一，小林隆夫，千葉健司，今井 聖，“メル一般化ケプストラム分析による音声のスペクトル推定”，電子情報通信学会論文誌.A, vol. J75-A, no. 7, pp. 1124-1134, 1992

【0026】

また、非特許文献1に記載されているように、ささやき声と通常の発声との音響分析から、より精緻にスペクトル包絡を変化させてもよい。例えば、ささやき声では1.5kHz以下のパワーが通常の発声に比べて小さいため、1.5kHz以下のスペクトル包絡を恣意的に小さくする方法が考えられる。また、1.5kHz以下のフォルマント周波数が通常の発声に比べ高くなっていることも報告されているため、推定されたスペクトル包絡の1.5kHz以下のフォルマント周波数を恣意的に高くすることで、よりささやき声に近づく変換処理を実現することができる。

10

【0027】

引き続き、図2を参照して、第一実施形態の音響モデル適応方法の処理手続きを説明する。

【0028】

疑似音声信号記憶部22には、学習データの音声信号を上述の疑似音声信号生成装置1により生成されたささやき声の疑似音声信号が記憶されている。

【0029】

ステップS13において、疑似音響特徴量生成部13は、疑似音声信号記憶部22に記憶された疑似音声信号を読み込み、疑似音声信号の各フレームから音響特徴量を抽出し、疑似音響特徴量系列を生成する。生成した疑似音響特徴量系列は、疑似音響特徴量記憶部23へ記憶される。抽出する音響特徴量としては、例えば、音声信号の短時間フレーム分析に基づくMFCCの1~12次元と、その動的特徴量である MFCC、 MFCCなどの動パラメータや、パワー、 パワー、 パワー等を用いる。また、MFCCに対してはケプストラム平均正規化(CMN: Cepstral Mean Normalization)処理を行ってもよい。抽出する音響特徴量は、MFCCやパワーに限定したものではなく、音声認識に用いられるパラメータを用いてもよい。

20

【0030】

ステップS14において、音響モデル学習部14は、疑似音響特徴量記憶部23に記憶された疑似音響特徴量系列を利用して音響モデルを学習する。学習済みの音響モデルは、音響モデル記憶部24へ記憶される。音声認識における音響モデルとしては、GMM-HMMなどが用いられており、音響モデルを認識対象タスクに適応させる手法は、例えば、下記参考文献2などに記載されている。

30

〔参考文献2〕篠田浩一，“確率モデルによる音声認識のための話者適応化技術”，電子情報通信学会論文誌、J87-D-II(2)、pp. 371-386、2004年

【0031】

音響モデルの適応に用いる音響特徴量は、音響モデルを用いる音声認識装置に求められる機能によって選択するとよい。具体的には、以下の2パターンが考えられる。

40

【0032】

(1) ささやき声に関する音響特徴量のみを用いて適応した音響モデルにより構築される音声認識装置の場合は、ささやき声の少量の学習データ(B)と疑似学習データ(C)のみを用いる。つまり、ささやき声に関する音響特徴量系列と疑似音響特徴量系列で適応した音響モデルを生成する。認識時には、ささやき声で発声した発話にのみ、この音響モデルを利用することができる。したがって、予めささやき声しか入力されないことがわかっている場合は、この音響モデルだけで音声認識装置を構築する。

【0033】

通常の発声による発話も含まれる場合は、通常の発声による十分な量の学習データ(A)のみで適応した音響モデルも併用すればよい。この場合、入力音声やささやき声である

50

か否かを判別して、二つの音響モデルのうちどちらを利用するかを判別するとよい。すなわち、認識対象の入力音声やささやき声である場合は、ささやき声に関するデータだけで適応された音響モデルを利用し、ささやき声でない場合は、通常の発声による音響特徴量だけで適応した音響モデルを利用する。なお、後述の第二実施形態で説明するボーカルフライに関する音響特徴量だけで適応した音響モデルも所有している場合には、さらにこの音響モデルも併用してよい。

**【 0 0 3 4 】**

ささやき声であるか否かを判別する方法は、例えば、(1) 音声の収録時に発話者がささやき声か否かを予め指定する方法、(2) 人間が実際に音声を聴取しささやき声か否かを判断する方法、(3) 自動でささやき声か否かを判別する方法が挙げられる。(1) 発話者が指定する方法は、音声を収録する際に発話者がこれから発話する音声やささやき声か否かを予め指定する。(2) 人間が判断する方法は、収録済みの音声を発話者本人もしくはそれ以外の人間が音声を聴取しささやき声か否かを判断する。(3) 自動で判別する方法は、例えば、音声信号をケプストラム分析した上で、その高次成分の大きさを予め定めた閾値と比較することでささやき声か否かを判別する。ささやき声は通常の発声とは異なり、ホワイトノイズのような非周期的な駆動音源信号となるため、駆動音源信号成分と考えられるケプストラムの高次成分の値が大きい場合は周期的と捉えて通常の発声であると判別し、小さい場合は非周期的と捉えてささやき声であると判別する。他には、ささやき声か否かを判別するモデルを予め構築しておき、そのモデルに基づいてささやき声か否かを判別する方法でもよい。例えば、ささやき声と通常の発声をそれぞれ混合ガウス分布 (GMM: Gaussian Mixture Model) により予めモデル化しておき、そのモデルに基づいた尤度比較によりささやき声か通常の発声かを識別する方法や、ささやき声と通常の発声の二つのクラスによるディープニューラルネットワーク (DNN: Deep Neural Networks) により識別する方法等も考えられる。

10

20

**【 0 0 3 5 】**

(2) ささやき声だけでなくすべての発声を含めた音響特徴量を用いて適応した音響モデルにより構築される音声認識装置の場合は、ささやき声ではない十分な量の学習データ (A) とささやき声の少量の学習データ (B) と疑似学習データ (C) とをすべて用いる。つまり、ささやき声に関する音響特徴量系列と疑似音響特徴量系列だけでなく、それ以外の全発話の音響特徴量系列で適応した音響モデルを生成する。このとき、第二実施形態で生成されるボーカルフライの疑似音響特徴量も含めてもよい。この場合、認識時には、すべての発話を一様に音声認識装置に入力することになる。

30

**【 0 0 3 6 】**

上述のように構成することで、第一実施形態の疑似音声信号生成装置および方法は、認識対象であるささやき声の学習データが十分に入手できない場合であっても、ささやき声と通常の発声との音響特性の違いを表現した変換ルールに基づいて、ささやき声の音声信号を疑似的に生成することができる。したがって、第一実施形態の音響モデル適応装置および方法は、ささやき声の疑似音声信号から抽出した十分な量の音響特徴量系列を用いて音響モデルの適応を行うことで、ささやき声に頑健な音響モデルを作成することができ、この音響モデルを用いて音声認識をすることで認識率が向上する。

40

**【 0 0 3 7 】****[ 第二実施形態 ]**

第二実施形態では、ルールベースによりボーカルフライの音声信号を疑似生成し、その疑似音声信号を用いて音響モデルを適応する方法について説明する。変換ルールはボーカルフライの音響特性に基づいて事前に作成しておき、その変換ルールに基づいて通常の発声からボーカルフライに変換することで、ボーカルフライによる音声信号の疑似生成を行う。

**【 0 0 3 8 】**

以下、第二実施形態の音響特徴量変換方法を説明する。以下では、上述の第一実施形態との相違点を中心に説明する。

50

## 【0039】

第二実施形態の疑似音声信号生成部12は、音声信号記憶部21に記憶されたデジタルの入力音声信号を読み込み、予め定めた変換ルールに基づいて入力音声信号の音響特性を変換してボーカルフライの疑似音声信号を生成する。生成した疑似音声信号は、疑似音声信号記憶部22へ記憶される。

## 【0040】

変換ルールはボーカルフライと通常の発声との違いを表現するようなルールである。ボーカルフライはフレーズの末尾等の定常的な母音音声部分に含まれることが多く、非特許文献2で報告されているように基本周波数が通常の発声より低くなる。そのため、取り得る変換ルールとしては、例えば、通常の発声による音声に対し、フレーズ末尾の母音の基本周波数を低くすることが考えられる。また、音声が無音区間を埋め込んだルールでもよい。

10

## 【0041】

上述のように構成することで、第二実施形態の疑似音声信号生成装置および方法は、認識対象であるボーカルフライの学習データが十分に入手できない場合であっても、ボーカルフライと通常の発声との音響特性の違いを表現した変換ルールに基づいて、ボーカルフライの音声信号を疑似的に生成することができる。したがって、第二実施形態の音響モデル適応装置および方法は、ボーカルフライの疑似音声信号から抽出した十分な量の音響特徴量系列を用いて音響モデルの適応を行うことで、ボーカルフライに頑健な音響モデルを作成することができ、この音響モデルを用いて音声認識をすることで認識率が向上する。

20

## 【0042】

この発明は上述の実施形態に限定されるものではなく、この発明の趣旨を逸脱しない範囲で適宜変更が可能であることはいうまでもない。上記実施形態において説明した各種の処理は、記載の順に従って時系列に実行されるのみならず、処理を実行する装置の処理能力あるいは必要に応じて並列的あるいは個別に実行されてもよい。

## 【0043】

[プログラム、記録媒体]

上記実施形態で説明した各装置における各種の処理機能をコンピュータによって実現する場合、各装置が有すべき機能の処理内容はプログラムによって記述される。そして、このプログラムをコンピュータで実行することにより、上記各装置における各種の処理機能がコンピュータ上で実現される。

30

## 【0044】

この処理内容を記述したプログラムは、コンピュータで読み取り可能な記録媒体に記録しておくことができる。コンピュータで読み取り可能な記録媒体としては、例えば、磁気記録装置、光ディスク、光磁気記録媒体、半導体メモリ等のようなものでもよい。

## 【0045】

また、このプログラムの流通は、例えば、そのプログラムを記録したDVD、CD-ROM等の可搬型記録媒体を販売、譲渡、貸与等することによって行う。さらに、このプログラムをサーバコンピュータの記憶装置に格納しておき、ネットワークを介して、サーバコンピュータから他のコンピュータにそのプログラムを転送することにより、このプログラムを流通させる構成としてもよい。

40

## 【0046】

このようなプログラムを実行するコンピュータは、例えば、まず、可搬型記録媒体に記録されたプログラムもしくはサーバコンピュータから転送されたプログラムを、一旦、自己の記憶装置に格納する。そして、処理の実行時、このコンピュータは、自己の記録媒体に格納されたプログラムを読み取り、読み取ったプログラムに従った処理を実行する。また、このプログラムの別の実行形態として、コンピュータが可搬型記録媒体から直接プログラムを読み取り、そのプログラムに従った処理を実行することとしてもよく、さらに、

50



このコンピュータにサーバコンピュータからプログラムが転送されるたびに、逐次、受け取ったプログラムに従った処理を実行することとしてもよい。また、サーバコンピュータから、このコンピュータへのプログラムの転送は行わず、その実行指示と結果取得のみによって処理機能を実現する、いわゆるASP (Application Service Provider) 型のサービスによって、上述の処理を実行する構成としてもよい。なお、本形態におけるプログラムには、電子計算機による処理の用に供する情報であってプログラムに準ずるもの(コンピュータに対する直接の指令ではないがコンピュータの処理を規定する性質を有するデータ等)を含むものとする。

【0047】

また、この形態では、コンピュータ上で所定のプログラムを実行させることにより、本装置を構成することとしたが、これらの処理内容の少なくとも一部をハードウェア的に実現することとしてもよい。

10

【符号の説明】

【0048】

- 1 疑似音声信号生成装置
- 2 音響モデル適応装置
- 11 音声信号取得部
- 12 疑似音声信号生成部
- 13 疑似音響特徴量生成部
- 14 音響モデル学習部
- 21 音声信号記憶部
- 22 疑似音声信号記憶部
- 23 疑似音響特徴量記憶部
- 24 音響モデル記憶部

20

【図1】

【図2】

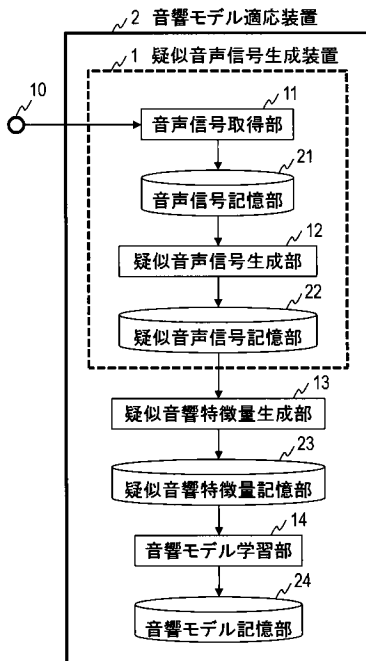


図1

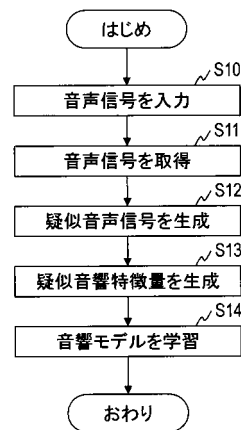


図2

フロントページの続き

(72)発明者 井島 勇祐

東京都千代田区大手町一丁目5番1号 日本電信電話株式会社内