

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2010 年 10 月 14 日(14.10.2010)



PCT



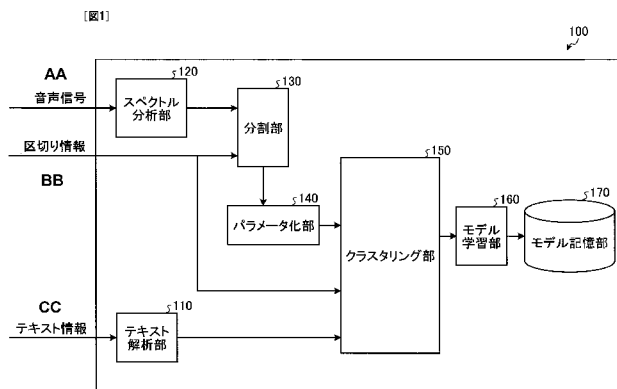
(10) 国際公開番号
WO 2010/116549 A1

- (51) 国際特許分類:
G10L 13/06 (2006.01)
- (21) 国際出願番号: PCT/JP2009/067408
- (22) 国際出願日: 2009 年 10 月 6 日(06.10.2009)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:
特願 2009-083563 2009 年 3 月 30 日(30.03.2009) JP
- (71) 出願人 (米国を除く全ての指定国について): 株式会社東芝(Kabushiki Kaisha Toshiba) [JP/JP]; 〒1058001 東京都港区芝浦一丁目 1 番 1 号 Tokyo (JP).
- (72) 発明者; および
- (75) 発明者/出願人 (米国についてののみ): ラトレ ハビエル(LATORRE, Javier) [ES/JP]; 〒1058001 東京都港区芝浦一丁目 1 番 1 号 株式会社東芝 知的財産部内 Tokyo (JP). 赤嶺 政巳(AKAMINE, Masami) [JP/JP]; 〒1058001 東京都港区芝浦一丁目 1 番 1 号 株式会社東芝 知的財産部内 Tokyo (JP).
- (74) 代理人: 酒井 宏明, 外(SAKAI, Hiroaki et al.); 〒1006020 東京都千代田区霞が関三丁目 2 番 5 号 霞が関ビルディング 酒井国際特許事務所 Tokyo (JP).
- (81) 指定国 (表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) 指定国 (表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ,

[続葉有]

(54) Title: SOUND MODEL GENERATION APPARATUS, SOUND SYNTHESIS APPARATUS, SOUND MODEL GENERATION PROGRAM, SOUND SYNTHESIS PROGRAM, SOUND MODEL GENERATION METHOD, AND SOUND SYNTHESIS METHOD

(54) 発明の名称: 音声モデル生成装置、音声合成装置、音声モデル生成プログラム、音声合成プログラム、音声モデル生成方法および音声合成方法



AA SOUND SIGNAL
BB BREAK INFORMATION
CC TEXT INFORMATION
120 SPECTRAL ANALYSIS UNIT
110 TEXT ANALYSIS UNIT
130 SPLITTING UNIT
140 PARAMETERIZATION UNIT
150 CLUSTERING UNIT
160 MODEL LEARNING UNIT
170 MODEL STORAGE UNIT

(57) Abstract: A learning model generation apparatus (100) is provided with a first calculation unit (120) for calculating, from each frame of a sound signal, a feature parameter indicating the spectral shape of the frame, a splitting unit (130) for splitting the sound signal into language sections, each having multiple frames and based on each language level, a parameterization unit (140) for, on the basis of the feature parameter of each of the multiple frames included in the language section, calculating the spectral parameter of the language section, a clustering unit (150) for clustering multiple spectral parameters calculated for multiple language sections, respectively, into multiple clusters on the basis of language information, and a model learning unit (160) for learning a spectral model indicating the features of the multiple spectral parameters from multiple spectral parameters included in the same cluster.

(57) 要約: 学習モデル生成装置 100 は、音声信号の各フレームから当該フレームのスペクトル形状を表す特徴パラメータを算出する第 1 算出部 120 と、音声信号を複数のフレームを有し、言語レベルを単位とする区間である言語区間に分割する分割部 130 と、言語区間に含まれる複数のフレームそれぞれの特徴パラメータに基づいて、言語区間のスペクトルパラメータを算出するパラメータ化部 140 と、複数の言語区間それぞれに対して算出された複数のスペクトルパラメータを、言語情報に基づいて複数のクラスターにクラスタリングするクラスタリング部 150 と、同一のクラスターに属する複数のスペクトルパラメータから複数のスペクトルパラメータの特徴を示すスペクトルモデルを学習するモデル学習部 160 とを備える。

WO 2010/116549 A1

WO 2010/116549 A1



CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, 添付公開書類:
TD, TG).

— 国際調査報告 (条約第 21 条(3))

明 細 書

発明の名称：

音声モデル生成装置、音声合成装置、音声モデル生成プログラム、音声合成プログラム、音声モデル生成方法および音声合成方法

技術分野

[0001] 本発明は、音声モデルを生成する音声モデル生成装置、音声モデルを用いて音声を合成する音声合成装置、音声モデル生成プログラム、音声合成プログラム、音声モデル生成方法および音声合成方法に関する。

背景技術

[0002] テキストから音声を生成する音声合成装置は、大別すると、テキスト解析部、韻律生成部及び音声信号生成部の3つの処理部から構成される。テキスト解析部では、言語辞書などを用いて入力されたテキスト（漢字かな混じり文）を解析し、漢字の読みやアクセントの位置、文節（アクセントの句）の区切りなどを定義した言語情報を出力する。韻律生成部では、言語情報に基づいて、声の高さ（基本周波数）の時間変化パターン（ピッチ包絡）と、各音韻の長さなどの音韻・韻律情報を出力する。音声信号生成部は、テキスト解析部からの音韻の系列と韻律生成部からの韻律情報に従って音声波形を生成するものであり、素片接続型合成方式とHMM合成方式の2方式が現在、主流となっている。

[0003] 素片接続型合成方式では、音韻の系列に従って音声素片を選択し、韻律情報に従って音声素片のピッチと継続時間長を変形して接続することで、合成音声を出力する。この方式は録音した音声データの素片を接続して音声波形を作成しているため比較的自然的な音質の合成音を得られる利点がある。しかしながら、素片を蓄積するためのメモリサイズが大きくなるという問題がある。

[0004] HMM合成方式は、合成フィルタをパルス列または雑音で駆動するボコーダーと呼ばれる合成器に基づいて合成音声を生成するものであり、統計モデ

ルに基づく音声合成方式の一つである。この方式では、合成器のパラメータを統計モデルで表現し、入力された文章に対して統計モデルの尤度が最大となるように合成器のパラメータを生成する。合成器のパラメータは、音声信号のスペクトルを表すLSFやFMCCなど、合成フィルタのパラメータと駆動信号のパラメータであり、それらの時系列は音素毎にHMMとガウス分布により統計的にモデル化される。学習用の音声データが与えられれば、統計モデルは音声データから自動的に学習することができ、メモリサイズも比較的小さくできる利点がある。

[0005] しかしながら、従来のHMM統計モデルに基づく音声合成方式では、スペクトルが統計的なモデル化により平均化されるため、生成される合成音の音質はメリハリのない籠った音質となるという問題がある。また、音素間でパラメータが不連続になり易く、異音が発生するという問題がある。

[0006] このようなパラメータの平均化や平滑化による音質の悪化を改善する方法として、文章全体にわたるスペクトルパラメータの分散を学習データから学習し、合成時に学習された分散を制約条件としてパラメータを生成、ダイナミクスを再生する手法が提案されている（非特許文献1）。

先行技術文献

非特許文献

[0007] 非特許文献1: Toda, T. and Tokuda K., 2005
“Speech Parameter Generation Algorithm Considering Global Variance for HMM-Based Speech Synthesis”. Proc. Interspeech 2005, Lisbon, Portugal, pp. 2801-2804

発明の概要

発明が解決しようとする課題

[0008] しかしながら、非特許文献1に記載されている方法は、スペクトルのメリ

ハリを回復させる効果があるものの、MFCCパラメータとの組み合わせ以外においては効果が確認されておらず、生成される合成フィルタがしばしば不安定なフィルタとなって異音が発生するという問題がある。

[0009] 本発明は、上記に鑑みてなされたものであって、滑らかに変化する自然なスペクトルを生成することのできる音声モデルを生成する音声モデル生成装置、この音声モデルを用いた音声合成装置、プログラムおよび方法を提供することを目的とする。

課題を解決するための手段

[0010] 上述した課題を解決し、目的を達成するために、本発明の一形態は、音声モデル生成装置に係り、テキスト情報を取得し、前記テキスト情報をテキスト解析することにより、前記テキスト情報に含まれる言語の内容を示す言語情報を生成するテキスト解析部と、前記テキスト情報に対応する音声信号を取得し、前記音声信号の各フレームから当該フレームのスペクトル形状を表す特徴パラメータを算出するスペクトル分析部と、前記音声信号の複数フレームを有し、言語レベルを単位とする区間である言語区間の境界位置を示す区切り情報を取得し、前記区切り情報に基づいて、前記音声信号を前記言語区間に分割する分割部と、前記言語区間に含まれる前記複数フレームそれぞれの前記特徴パラメータに基づいて、前記言語区間のスペクトルパラメータを算出するパラメータ化部と、複数の言語区間それぞれに対して算出された複数のスペクトルパラメータを、前記言語情報に基づいて複数のクラスターにクラスタリングするクラスタリング部と、同一のクラスターに属する複数のスペクトルパラメータから前記複数のスペクトルパラメータの特徴を示すスペクトルモデルを学習するモデル学習部とを備えることを特徴とする。

[0011] また、本発明の他の形態は、音声合成装置に係り、音声合成の対象となるテキスト情報を取得し、前記テキスト情報をテキスト解析することにより、前記テキスト情報に含まれる言語の内容を示す言語情報を生成するテキスト解析部と、複数フレームを有する言語レベルを単位とする言語区間に含まれる、前記テキスト情報に対応する複数の音声信号それぞれの複数のスペクト

ルパラメータの特徴を示すスペクトルモデルであって、前記言語区間の前記言語情報により複数のクラスターにクラスタリングされたスペクトルモデルを記憶する記憶部から、前記音声合成の対象となるテキスト情報の前記言語区間の前記言語情報に基づいて、前記テキスト情報の前記言語区間が属する前記クラスターの前記スペクトルモデルを選択する選択部と、前記選択部により選択された前記スペクトルモデルに基づいて、前記言語区間に対するスペクトルパラメータを生成し、前記スペクトルパラメータを逆変換することにより、特徴パラメータを得る生成部とを備えることを特徴とする。

発明の効果

- [0012] 本発明によれば、複数フレームを含む言語区間単位でスペクトルモデルを学習するので、このスペクトルモデルを用いて音声合成を行うことにより、不連続点のない自然なスペクトルを得ることができるという効果を奏する。

図面の簡単な説明

- [0013] [図1]学習モデル生成装置 100の構成を示すブロック図である。
[図2]言語区間を説明するための図である。
[図3]決定木の一例を示す図である。
[図4]学習モデル生成処理を示すフローチャートである。
[図5]パラメータ化部 140により得られたスペクトルパラメータを示す図である。
[図6]HMMによりフレーム単位で得られたスペクトルパラメータを示す図である。
[図7]音声合成装置 200の構成を示す図である。
[図8]音声合成装置 200による音声合成処理を示すフローチャートである。
[図9]学習モデル生成装置 100のハードウェア構成を示す図である。

発明を実施するための形態

- [0014] 以下に添付図面を参照して、この発明にかかる音声モデル生成装置、音声合成装置、プログラムおよび方法の最良な実施の形態を詳細に説明する。
- [0015] 図1は、本発明の実施の形態にかかる学習モデル生成装置 100の構成を

示すブロック図である。学習モデル生成装置 100 は、テキスト解析部 110 と、スペクトル分析部 120 と、分割部 130 と、パラメータ化部 140 と、クラスタリング部 150 と、モデル学習部 160 と、モデル記憶部 170 とを備えている。学習モデル生成装置 100 は、テキスト情報と、テキスト情報の内容を読み上げた音声信号とを学習データとして取得し、学習データに基づいて、音声合成のための学習モデルを生成する。

[0016] テキスト解析部 110 は、テキスト情報を取得する。テキスト解析部 110 は、取得したテキスト情報に対するテキスト解析により言語情報を生成する。ここで、言語情報は、言語レベルを単位とする言語区間の境界位置を示す区間情報、各言語区間の形態素、各言語区間の音素記号、各音素が有声音であるか無声音であるかを示す情報、各音素のアクセントの有無を示す情報、各言語区間の開始時間、終了時間、各言語区間の前後の言語区間の情報、各言語区間と前後の言語区間との言語的な関係を示す情報など言語の内容を示す情報である。言語情報はコンテキストと呼ばれ、クラスタリング部 150 において、スペクトルパラメータのコンテキストモデル作成に用いられる。なお、言語区間とは、複数フレームを含み、所定の言語レベルを単位とする区間である。言語レベルとしては、音素、音節、単語、句、呼気段階、発声全体などがある。

[0017] スペクトル分析部 120 は、音声信号を取得する。音声信号は、テキスト解析部 110 が取得したテキスト情報の内容を読み上げた発話についての音声の信号である。音声信号は、学習のための音声データを発話単位に分割したものである。

[0018] スペクトル分析部 120 は、取得した音声信号に対し、スペクトル分析を行う。すなわち、音声信号を 10ms 程度のフレームに分割する。そして、フレーム毎に、フレームのスペクトルの形状を表す特徴パラメータとしてのメルケプストラム係数 (MFCC) を算出し、各フレームの音声信号と MFCC の組を分割部 130 に出力する。

[0019] 分割部 130 は、外部から区切り情報を取得する。区切り情報とは、音声

信号の言語レベル単位での境界位置、すなわち言語区間の境界位置を示す情報である。区切り情報は、マニュアルまたは自動的なアライメントにより生成される。自動的なアライメントとしては、例えば、HMMで構成される音声認識モデルを用いて、入力された音声信号のフレームを音響モデルの状態に対応付け、この対応付けから言語区間の区切り情報を得る。区切り情報は、学習データとともに与えられるものとする。分割部130は、区切り情報に基づいて、音声信号の言語区間を特定し、スペクトル分析部120から取得したMFCCを言語区間に分割する。

[0020] 図2に示すように、例えば[k a i r o]というテキスト情報に対応するMFCC曲線は、音素単位では、／k／、／a i／、／r／、／o／の4つの音素の言語区間に区切られる。分割部130は、例えば音素、音節、単語、句、呼気段階および発声全体など複数の言語レベルにおいてMFCCを言語区間に分割する。

[0021] なお、これ以降で説明する処理においても、各言語レベルの言語区間それぞれに対して処理が施されるが、以下の説明においては、一例として、音素を言語レベルとする場合について述べる。

[0022] パラメータ化部140は、MFCCを分割部130において区切られた単位、すなわち言語区間単位でベクトルとし、そのベクトルからスペクトルパラメータを算出する。なお、スペクトルパラメータは、基本パラメータと拡張パラメータとを有している。

[0023] パラメータ化部140は、言語区間に含まれるフレーム数を k とした場合、複数フレームのMFCCから構成される k 次元ベクトル $M e l C e p i, s$ に対し、(式1)に示すように、 k 次のDCTを適用することにより、基本パラメータを算出する。このように、基本パラメータは、対象とする言語区間である対象区間のスペクトルパラメータであり、対象区間の特徴を示すパラメータである。

[数1]

$$X_{i,s} = T_{i,s} \cdot \text{MelCep}_{i,s} \cdot \cdot \cdot \cdot \quad (\text{式1})$$

なお、 $\text{MelCep}_{i,s}$ は、音素 s の i 次の MFCC 係数の k 次元ベクトルである。 $T_{i,s}$ は、音素 s のフレーム数 k に対応する k 次の DCT の変換行列である。DCT の次元は言語レベルの単位やフレーム長などに依存する。なお、基本フレームを算出する際には、DCT 以外の種々の線形変換を用いてもよい。例えば、逆変換可能な離散コサイン変換、フーリエ変換、ウェーブレット変換、テーラー展開および多項式展開を用いてもよい。

[0024] パラメータ化部 140 は、さらに拡張パラメータを算出する。拡張パラメータは、対象区間に隣接する言語区間の MFCC ベクトルの傾きで構成される。なお、隣接する区間とは、対象区間の直前の言語区間である直前区間と、対象区間の直後の言語区間である直後区間である。直前区間の拡張パラメータ

[数2]

$$\Delta \text{MelCep}_{i,s}^{\text{left}}$$

および直後区間の拡張パラメータ

[数3]

$$\Delta \text{MelCep}_{i,s}^{\text{right}}$$

は、それぞれ（式2）および（式3）により算出される。ここで、 α は傾きを計算するための W 次元重みベクトルである。また、カッコ内の負のインデックスはベクトルの最後の要素から数えた場合の要素を示している。

[数4]

$$\Delta \text{MelCep}_{i,s}^{\text{left}} = \sum_{w=0}^W \alpha(w) \cdot \text{MelCep}_{i,s}(w) + \sum_{w=-W}^{-1} \alpha(w) \text{MelCep}_{i,s-1}(-w) \cdot \cdot \cdot \cdot \quad (\text{式2})$$

[数5]

$$\Delta \text{MelCep}_{i,s}^{\text{right}} = \sum_{w=-W}^0 \alpha(w) \cdot \text{MelCep}_{i,s}(w) + \sum_{w=1}^W \alpha(w) \cdot \text{MelCep}_{i,s+1}(w) \cdot \cdot \cdot \cdot \quad (\text{式3})$$

[0025] 上記の拡張パラメータは、基本パラメータを用いて、それぞれ（式4）、（式5）のように書き換えることができる。すなわち、拡張パラメータを基本パラメータの関数として表すことができる。

[数6]

$$\Delta \text{MelCep}_{i,s}^{\text{left}} = H_{i,s}^{\text{begin}} \cdot X_{i,s} + H_{i,s-1}^{\text{end}} \cdot X_{i,s-1} \cdot \cdot \cdot \cdot \quad (\text{式4})$$

[数7]

$$\Delta \text{MelCep}_{i,s}^{\text{right}} = H_{i,s}^{\text{end}} \cdot X_{i,s} + H_{s+1}^{\text{begin}} \cdot X_{i,s-1} \cdot \cdot \cdot \cdot \quad (\text{式5})$$

なお、

[数8]

$$H_{i,s}^{\text{begin}}$$

および

[数9]

$$H_{i,s}^{\text{end}}$$

は、それぞれ、（式6）、（式7）で表される。

[数10]

$$H_{i,s}^{\text{begin}} = \sum_{w=0}^W \alpha(w) \cdot T_{i,s}^{-1}(w) \cdot \cdot \cdot \cdot \quad (\text{式6})$$

[数11]

$$H_{i,s}^{\text{end}} = \sum_{w=-W}^0 \alpha(w) \cdot T_{i,s}^{-1}(-w) \cdots \quad (\text{式7})$$

[0026] パラメータ化部 140 は、分割部 130 により算出された基本パラメータおよび拡張パラメータを（式 8）に示すように、1つのスペクトルパラメータ $SP_{i,s}$ に統合する。

[数12]

$$SP_{i,s} = (X_{i,s}^t, \Delta \text{MelCep}_{i,s}^{\text{left}}, \Delta \text{MelCep}_{i,s}^{\text{right}})^t \cdots \quad (\text{式8})$$

[0027] クラスタリング部 150 は、パラメータ化部 140 により得られた各言語区間のスペクトルパラメータを、区切り情報およびテキスト解析部 110 により生成された言語情報に基づいてクラスタリングする。具体的には、クラスタリング部 150 は、言語情報、すなわちコンテキスト情報に関する質問を繰り返しながら分岐を繰り返す決定木に基づいて、スペクトルパラメータを複数のクラスターに分割する。例えば、図 3 に示すように、「対象区間は / a / か？」といった質問に対する Yes、No の答えに応じてスペクトルパラメータは Yes の子ノードと No の子ノードに分割される。質問と、回答によるスペクトルパラメータの分割が繰り返されて、図 3 に示すように言語情報に関する条件が等しい複数のスペクトルパラメータが同一クラスターにグループ化される。

[0028] 図 3 に示す例においては、対象区間、直前区間および直後区間の音素が等しい対象区間のスペクトルパラメータが同一のクラスターになるように分類されている。図 3 に示す例においては、対象区間としての音素 / a / であっても、直前直後の音素が異なる [(k) a (n)] と、[(k) a (m)] はそれぞれ異なるクラスターに分類される。

[0029] なお、上記において説明したクラスターは一例であり、他の例としては、上述のように、対象区間、直前区間および直後区間の音素のほか、対象区間

におけるアクセントの有無、直前区間、直後区間におけるアクセントの有無など、各区間の音素以外の言語情報を用いてより細かいクラスターに分類してもよい。

[0030] また、クラスタリングはMFCCの全次元の係数ベクトルに対応する基本パラメータと拡張パラメータを統合したスペクトルパラメータに対して行うこととしたが、他の例としては、MFCCの次元ごとに行ってもよい。各次元でクラスタリングする場合は、クラスタリングするスペクトルパラメータの次元が統合したスペクトルパラメータの次元より小さくなる。このため、クラスタリングの精度を向上させることができる。同様に、統合したスペクトルパラメータの次元をPCA (Principal Component Analysis: 主成分分析) の手法を用いて次元圧縮した後に行ってもよい。

[0031] モデル学習部160は、各クラスターに分類された複数のスペクトルパラメータから、これら複数のスペクトルパラメータの分布を近似するガウス分布のパラメータを学習し、コンテキスト依存のスペクトルモデルとして出力する。具体的には、モデル学習部160は、 $SP_{mi, s}$ 、平均ベクトル $m_{i, s}$ および共分散行列 $\Sigma_{i, s}$ の3つのパラメータをスペクトルモデルとして出力する。なお、クラスタリングの方法やガウス分布のパラメータ学習法としては、音声認識の分野でよく知られている方法を利用することができる。

[0032] モデル記憶部170は、モデル学習部160により出力された学習モデルを、学習モデルに共通する言語情報の条件に対応付けて記憶する。なお、言語情報の条件とは、クラスタリングにおいて質問に用いた言語情報である。

[0033] 図4は、学習モデル生成装置100による学習モデル生成処理を示すフローチャートである。学習モデル生成処理においては、まず学習モデル生成装置100は学習データとしてテキスト情報、テキストの区切り位置を示す区切り情報およびテキストに対応する音声信号を取得する(ステップS100)。具体的には、テキスト情報はテキスト解析部110、音声信号はスペク

トル分析部 120、区切り情報は、分割部 130 およびクラスタリング部 150 に入力される。

[0034] 次に、テキスト解析部 110 は、テキスト情報に基づいて、言語情報を生成する（ステップ S102）。スペクトル分析部 120 は、音声信号の各フレームの特徴パラメータ MFCC を算出する（ステップ S104）。なお、テキスト解析部 110 による言語情報の生成およびスペクトル分析部 120 による特徴パラメータ算出の処理は独立に行われるので、両者の処理順番は問わない。

[0035] 次に、分割部 130 は、区切り情報に基づいて、音声信号の言語区間を特定する（ステップ S106）。次に、パラメータ化部 140 は、言語区間に含まれる複数のフレームそれぞれの MFCC から言語区間のスペクトルパラメータを算出する（ステップ S108）。パラメータ化部 140 はより詳しくは、対象区間だけでなく、対象区間の直前区間、直後区間それぞれに含まれる複数フレームの MFCC に基づいて、基本パラメータおよび拡張パラメータを要素とするスペクトルパラメータ $SP_{i,s}$ を算出する。

[0036] 次に、クラスタリング部 150 は、パラメータ化部 140 によりテキスト情報の各言語区間に対して得られた複数のスペクトルパラメータを、区切り情報および言語情報に基づいてクラスタリングする（ステップ S110）。次に、モデル学習部 160 は、各クラスターに属する複数のスペクトルパラメータから学習モデルとしてのスペクトルモデルを生成する（ステップ S112）。次に、モデル学習部 160 は、スペクトルモデルを、対応するテキスト情報および言語情報（言語情報の条件）に対応付けてモデル記憶部 170 に記憶する（ステップ S114）。以上で、学習モデル生成装置 100 による学習モデル生成処理が完了する。

[0037] 図 5 および図 6 からわかるように、本実施の形態にかかる学習モデル生成装置 100 は、HMM によるスペクトルパラメータに比べて、より実際のスペクトルに近いスペクトルパラメータを得ることができる。学習モデル生成装置 100 は、複数フレームに対応する言語区間を単位とするスペクトルパ

ラメータからスペクトルモデルを学習するので、より自然なスペクトルモデルを得ることができる。さらに、このスペクトルモデルを利用することにより、より自然なスペクトルパターンを生成することができる。

[0038] また、学習モデル生成装置 100 は、対象区間に対応する基本パラメータだけでなく、直前区間および直後区間に対応する拡張パラメータを考慮することにより、不連続点が生じることなく滑らかに変化するスペクトルモデルを学習することができる。

[0039] さらに、学習モデル生成装置 100 は、複数の言語レベルそれぞれに対するスペクトルモデルを学習するので、これらのスペクトルモデルを利用して、総合的なスペクトルパターンを生成することができる。

[0040] 図 7 は、音声合成装置 200 の構成を示す図である。音声合成装置 200 は、音声合成の対象となるテキスト情報を取得し、学習モデル生成装置 100 により生成されたスペクトルモデルに基づいて、音声合成を行う。音声合成装置 200 は、モデル記憶部 210 と、テキスト解析部 220 と、モデル選択部 230 と、継続時間長算出部 240 と、スペクトルパラメータ生成部 250 と、F0 生成部 260 と、駆動信号生成部 270 と、合成フィルタ 280 とを備えている。

[0041] モデル記憶部 210 は、学習モデル生成装置 100 において生成された学習モデルを言語情報の条件に対応付けて記憶している。なお、モデル記憶部 210 は、学習モデル生成装置 100 のモデル記憶部 170 と同様である。テキスト解析部 220 は、外部から音声合成の対象となるテキスト情報を取得する。テキスト解析部 220 は、テキスト情報に対し、テキスト解析部 110 と同様の処理を行う。すなわち、取得したテキスト情報に対応する言語情報を生成する。モデル選択部 230 は、言語情報に基づいて、テキスト解析部 220 に入力されたテキスト情報に含まれる複数の言語区間それぞれに対応する、コンテキスト依存のスペクトルモデルをモデル記憶部 210 から選択する。モデル選択部 230 は、テキスト情報に含まれる複数の言語区間それぞれに対して選択されたスペクトルモデルを接続し、これをテキスト情

報全体に対応するモデル系列として出力する。

[0042] 継続時間長算出部 240 は、テキスト解析部 220 から言語情報を取得し、言語情報に定義された各言語区間の開始時間と終了時間とに基づいて、各言語区間の継続時間長を算出する。

[0043] スペクトルパラメータ生成部 250 は、モデル選択部 230 により選択された言語区間のモデル系列と、継続時間長算出部 240 により各言語区間に対して算出された継続時間長を接続した継続時間長系列とを入力とし、入力されたテキスト全体に対応するスペクトルパラメータを算出する。具体的には、モデル系列と継続時間長系列とに基づいて、スペクトルパラメータ $SP_{i,s}$ の対数尤度（尤度関数）を総目的関数 F とし、目的関数が最大となるようなスペクトルパラメータを算出する。総目的関数 F は、（式 9）で表される。

[数13]

$$F = \sum_{\forall s} \log(P(SP_{i,s} | s)) \cdots \cdots \text{（式9）}$$

ここで、 s は、単位区間の集合である。スペクトルパラメータはガウス分布でモデル化されているので、その確率は（式 10）に示すように、ガウス分布の確率密度で与えられる。

[数14]

$$P(SP_{i,s} | s) = N(SP_{i,s}; \mu_{i,s}, \Sigma_{i,s}) \cdots \cdots \text{（式10）}$$

[0044] スペクトルパラメータを求めるべく、この総目的関数 F を基準となる言語レベル（音素）でのスペクトルパラメータ $X_{i,s}$ について最大化する。パラメータの最大化は、勾配法などの公知の技術を用いるものとする。このように、目的関数を最大化することにより、適切なスペクトルパラメータを算出することができる。

[0045] 他の例としては、スペクトルパラメータ生成部 250 は、スペクトルのグ

ローバル分散も考慮に入れて目的関数を最大化することとしてもよい。これにより、生成されるスペクトルのパターンが自然音声のスペクトルパターンの変化幅と同様に变化し、より自然な音声を得ることができる。

[0046] スペクトルパラメータ生成部 250 は、目的関数の最大化で導出されたスペクトルの基本パラメータ α_i , s を逆変換することで、音素に含まれる複数フレームの MFCC 係数を生成する。なお、逆変換は、言語区間に含まれる複数のフレームに渡って行う。

[0047] F0 生成部 260 は、テキスト解析部 220 から言語情報を取得し、継続時間長算出部 240 から各言語区間の継続時間長を取得する。F0 生成部 260 は、言語情報に含まれるアクセントの有無などの情報および各言語区間の継続時間長に基づいて、ピッチの基本周波数 (F0) を生成する。

[0048] 駆動信号生成部 270 は、F0 生成部 260 から基本周波数 (F0) を取得し、基本周波数 (F0) から駆動信号を生成する。具体的には、対象区間が有声音である場合には、基本周波数 (F0) の逆数であるピッチ周期のパルス列を駆動信号として生成する。また、対象区間が無声音である場合、白色雑音を駆動信号として生成する。

[0049] 合成フィルタ 280 は、スペクトルパラメータ生成部 250 により得られたスペクトルパラメータおよび駆動信号生成部 270 により生成された駆動信号から合成フィルタを用いて合成音声を生じ出力する。具体的には、まずスペクトルパラメータである MFCC パラメータを LPC パラメータに変換する。そして、LPC パラメータを有する全極フィルタを適用する。LPC パラメータを α_i ($i = 1, 2, 3 \dots, p$) とした場合、合成フィルタとしての全極フィルタの伝達関数 $H(z)$ は、(式 11) で表される。ここで、 p は合成フィルタの次数である。

[数15]

$$H(Z) = 1 / \sum_{i=1}^p \alpha_i Z^{-i} \dots \dots (式11)$$

[0050] また、全極フィルタへの入力信号である駆動信号を $e(n)$ 、全極フィルタの出力を $y(n)$ とした場合、合成フィルタの動作は（式 12）の差分方程式で表される。

[数16]

$$y(n) = \sum_{i=1}^p \alpha_i y(n-i) + e(n) \cdots \cdots \text{（式12）}$$

[0051] 図 8 は、音声合成装置 200 による音声合成処理を示すフローチャートである。音声合成処理において、まずテキスト解析部 220 は音声合成の対象となるテキスト情報を取得する（ステップ S 200）。次に、テキスト解析部 220 は、取得したテキスト情報に基づいて、言語情報を生成する（ステップ S 202）。次に、モデル選択部 230 は、テキスト解析部 220 が生成した言語情報に基づいて、モデル記憶部 210 からテキスト情報に含まれる各言語区間に対するスペクトルモデルを選択し、これらを接続したモデル系列を得る（ステップ S 204）。次に、継続時間長算出部 240 は、言語情報に含まれる各言語区間の開始時間および終了時間に基づいて、各言語区間の継続時間長を算出する（ステップ S 206）。なお、モデル選択部 230 によるモデル選択処理および継続時間長算出部 240 による継続時間長算出処理は独立した処理であり、これらの処理順番は特に限定されるものではない。

[0052] 次に、スペクトルパラメータ生成部 250 は、モデル系列および継続時間長系列に基づいて、テキスト情報に対応するスペクトルパラメータを算出する（ステップ S 208）。次に、F0 生成部 260 は、言語情報および継続時間長に基づいて、ピッチの基本周波数（F0）を生成する（ステップ S 210）。次に、駆動信号生成部 270 は、駆動信号を生成する（ステップ S 212）。次に、合成フィルタ 280 により合成音声信号が生成され外部に出力されて（ステップ S 214）、音声合成処理が完了する。

[0053] このように、本実施の形態にかかる音声合成装置 200 は、学習モデル生

成装置 100 により生成された、DCT 係数で表現されたスペクトルモデルを利用して音声合成を行うので、滑らかに変化する自然なスペクトルを生成することができる。

- [0054] 図 9 は、学習モデル生成装置 100 のハードウェア構成を示す図である。学習モデル生成装置 100 は、CPU (Central Processing Unit) 11 と、ROM (Read Only Memory) 12 と、RAM (Random Access Memory) 13 と、記憶部 14 と、表示部 15 と、操作部 16 と、通信部 17 とを備え、各部はバス 18 を介して接続されている。
- [0055] CPU 11 は、RAM 13 を作業領域として、ROM 12 又は記憶部 14 に記憶されたプログラムとの協働により各種処理を実行し、学習モデル生成装置 100 の動作を統括的に制御する。また、CPU 11 は、ROM 12 又は記憶部 14 に記憶されたプログラムとの協働により、上述の各機能部を実現させる。
- [0056] ROM 12 は、学習モデル生成装置 100 の制御にかかるプログラムや各種設定情報などを書き換え不可能に記憶する。RAM 13 は、SDRAM や DDR メモリなどの揮発性メモリであって、CPU 11 の作業エリアとして機能する。
- [0057] 記憶部 14 は、磁氣的又は光学的に記録可能な記憶媒体を有し、学習モデル生成装置 100 の制御にかかるプログラムや各種情報を書き換え可能に記憶する。また、記憶部 14 は、上述のモデル学習部 160 により生成されるスペクトルモデルなどを記憶する。表示部 15 は、LCD (Liquid Crystal Display) などの表示デバイスから構成され、CPU 11 の制御の下、文字や画像などを表示する。操作部 16 は、マウスやキーボードなどの入力デバイスであって、ユーザから操作入力された情報を指示信号として受け付け、CPU 11 に出力する。通信部 17 は、外部装置との間で通信を行うインターフェイスであって、外部装置から受信した各種情報を CPU 11 に出力する。また、通信部 17 は、CPU 11 の制御の下、

各種情報を外部装置に送信する。なお、音声合成装置 200 のハードウェア構成は、学習モデル生成装置 100 のハードウェア構成と同様である。

[0058] 本実施の形態にかかる学習モデル生成装置 100 および音声合成装置 200 において実行される学習モデル生成プログラムおよび音声合成プログラムは、ROM等に予め組み込まれて提供される。

[0059] 本実施の形態の学習モデル生成装置 100 および音声合成装置 200 で実行される学習モデル生成プログラムおよび音声合成プログラムは、インストール可能な形式又は実行可能な形式のファイルで CD-ROM、フレキシブルディスク (FD)、CD-R、DVD (Digital Versatile Disk) 等のコンピュータで読み取り可能な記録媒体に記録して提供するように構成してもよい。

[0060] さらに、本実施の形態の学習モデル生成装置 100 および音声合成装置 200 で実行される学習モデル生成プログラムおよび音声合成プログラムを、インターネット等のネットワークに接続されたコンピュータ上に格納し、ネットワーク経由でダウンロードさせることにより提供するように構成しても良い。また、本実施形態の学習モデル生成装置 100 および音声合成装置 200 で実行される学習モデル生成プログラムおよび音声合成プログラムをインターネット等のネットワーク経由で提供または配布するように構成しても良い。

[0061] 本実施の形態の学習モデル生成装置 100 および音声合成装置 200 で実行される学習モデル生成プログラムおよび音声合成プログラムは、上述した各部を含むモジュール構成となっており、実際のハードウェアとしては CPU (プロセッサ) が上記 ROM から学習モデル生成プログラムおよび音声合成プログラムを読み出して実行することにより上記各部が主記憶装置上にロードされ、上述した各部が主記憶装置上に生成されるようになっている。

[0062] なお、本発明は、上記実施の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することができる。また、上記実施の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、種々の発明を形成することができる。例えば、実施の形態に示される全構成要素からいくつかの構成要素を削除してもよい。さらに、異なる実施の形態にわたる構成要素を適宜組み合わせても良い。

符号の説明

- [0063] 1 0 0 学習モデル生成装置
- 1 2 0 スペクトル分析部
- 1 3 0 分割部
- 1 4 0 パラメータ化部
- 1 5 0 クラスタリング部
- 1 6 0 モデル学習部

請求の範囲

- [請求項1] テキスト情報を取得し、前記テキスト情報をテキスト解析することにより、前記テキスト情報に含まれる言語の内容を示す言語情報を生成するテキスト解析部と、
- 前記テキスト情報に対応する音声信号を取得し、前記音声信号の各フレームから当該フレームのスペクトル形状を表す特徴パラメータを算出するスペクトル分析部と、
- 前記音声信号の複数フレームを有し、言語レベルを単位とする区間である言語区間の境界位置を示す区切り情報を取得し、前記区切り情報に基づいて、前記音声信号を前記言語区間に分割する分割部と、
- 前記言語区間に含まれる前記複数フレームそれぞれの前記特徴パラメータに基づいて、前記言語区間のスペクトルパラメータを算出するパラメータ化部と、
- 複数の言語区間それぞれに対して算出された複数のスペクトルパラメータを、前記言語情報に基づいて複数のクラスターにクラスタリングするクラスタリング部と、
- 同一のクラスターに属する複数のスペクトルパラメータから前記複数のスペクトルパラメータの特徴を示すスペクトルモデルを学習するモデル学習部と
- を備えることを特徴とする音声モデル生成装置。
- [請求項2] 前記パラメータ化部は、処理対象となる前記言語区間である対象区間に含まれる前記複数フレームそれぞれの前記特徴パラメータおよび前記対象区間の直前および直後の前記言語区間それぞれに含まれる前記複数フレームそれぞれの前記特徴パラメータに基づいて、前記対象区間の前記スペクトルパラメータを算出することを特徴とする請求項1に記載の音声モデル生成装置。
- [請求項3] 前記モデル学習部は、前記対象区間、前記対象区間の直前および直後の前記言語区間により、前記対象区間を複数のクラスターにクラス

タリングすることを特徴とする請求項 2 に記載の音声モデル生成装置。

[請求項 4] 前記パラメータ化部は、前記対象区間に含まれる複数の前記特徴パラメータに対し、線形変換を施すことにより、前記スペクトルパラメータを得ることを特徴とする請求項 1 に記載の音声モデル生成装置。

[請求項 5] 音声合成の対象となるテキスト情報を取得し、前記テキスト情報をテキスト解析することにより、前記テキスト情報に含まれる言語の内容を示す言語情報を生成するテキスト解析部と、

複数フレームを有する言語レベルを単位とする言語区間に含まれる、前記テキスト情報に対応する複数の音声信号それぞれの複数のスペクトルパラメータの特徴を示すスペクトルモデルであって、前記言語区間の前記言語情報により複数のクラスターにクラスタリングされたスペクトルモデルを記憶する記憶部から、前記音声合成の対象となるテキスト情報の前記言語区間の前記言語情報に基づいて、前記テキスト情報の前記言語区間が属する前記クラスターの前記スペクトルモデルを選択する選択部と、

前記選択部により選択された前記スペクトルモデルに基づいて、前記言語区間に対するスペクトルパラメータを生成し、前記スペクトルパラメータを逆変換することにより、特徴パラメータを得る生成部とを備えることを特徴とする音声合成装置。

[請求項 6] 前記生成部は、前記選択部により選択された前記スペクトルモデルの目的関数を生成し、前記目的関数を最大化することにより、前記言語区間に対するスペクトルパラメータを生成することを特徴とする請求項 5 に記載の音声合成装置。

[請求項 7] 音声モデル生成処理をコンピュータに実行させるための音声モデル生成プログラムであって、

前記コンピュータを、

テキスト情報を取得し、前記テキスト情報をテキスト解析すること

により、前記テキスト情報に含まれる言語の内容を示す言語情報を生成するテキスト解析部と、

前記テキスト情報に対応する音声信号を取得し、前記音声信号の各フレームから当該フレームのスペクトル形状を表す特徴パラメータを算出するスペクトル分析部と、

前記音声信号の複数フレームを有し、言語レベルを単位とする区間である言語区間の境界位置を示す区切り情報を取得し、前記区切り情報に基づいて、前記音声信号を前記言語区間に分割する分割部と、

前記言語区間に含まれる前記複数フレームそれぞれの前記特徴パラメータに基づいて、前記言語区間のスペクトルパラメータを算出するパラメータ化部と、

複数の言語区間それぞれに対して算出された複数のスペクトルパラメータを、前記言語情報に基づいて複数のクラスターにクラスタリングするクラスタリング部と、

同一のクラスターに属する複数のスペクトルパラメータから前記複数のスペクトルパラメータの特徴を示すスペクトルモデルを学習するモデル学習部と

して機能させるためのプログラム。

[請求項8] 音声合成処理をコンピュータに実行させるための音声合成プログラムであって、

前記コンピュータを、

音声合成の対象となるテキスト情報を取得し、前記テキスト情報をテキスト解析することにより、前記テキスト情報に含まれる言語の内容を示す言語情報を生成するテキスト解析部と、

複数フレームを有する言語レベルを単位とする言語区間に含まれる、前記テキスト情報に対応する複数の音声信号それぞれの複数のスペクトルパラメータの特徴を示すスペクトルモデルであって、前記言語区間の前記言語情報により複数のクラスターにクラスタリングされた

スペクトルモデルを記憶する記憶部から、前記音声合成の対象となるテキスト情報の前記言語区間の前記言語情報に基づいて、前記テキスト情報の前記言語区間が属する前記クラスターの前記スペクトルモデルを選択する選択部と、

前記選択部により選択された前記スペクトルモデルに基づいて、前記言語区間に対するスペクトルパラメータを生成し、前記スペクトルパラメータを逆変換することにより、特徴パラメータを得る生成部として機能させるためのプログラム。

[請求項9]

テキスト解析部が、テキスト情報を取得し、前記テキスト情報をテキスト解析することにより、前記テキスト情報に含まれる言語の内容を示す言語情報を生成するテキスト解析ステップと、

スペクトル分析部が、前記テキスト情報に対応する音声信号を取得し、前記音声信号の各フレームから当該フレームのスペクトル形状を表す特徴パラメータを算出するスペクトル分析ステップと、

分割部が、前記音声信号の複数フレームを有し、言語レベルを単位とする区間である言語区間の境界位置を示す区切り情報を取得し、前記区切り情報に基づいて、前記音声信号を前記言語区間に分割する分割ステップと、

パラメータ化部が、前記言語区間に含まれる前記複数フレームそれぞれの前記特徴パラメータに基づいて、前記言語区間のスペクトルパラメータを算出するパラメータ化ステップと、

クラスタリング部が、複数の言語区間それぞれに対して算出された複数のスペクトルパラメータを、前記言語情報に基づいて複数のクラスターにクラスタリングするクラスタリングステップと、

モデル学習部が、同一のクラスターに属する複数のスペクトルパラメータから前記複数のスペクトルパラメータの特徴を示すスペクトルモデルを学習する学習ステップと

を有することを特徴とする音声モデル生成方法。

[請求項10]

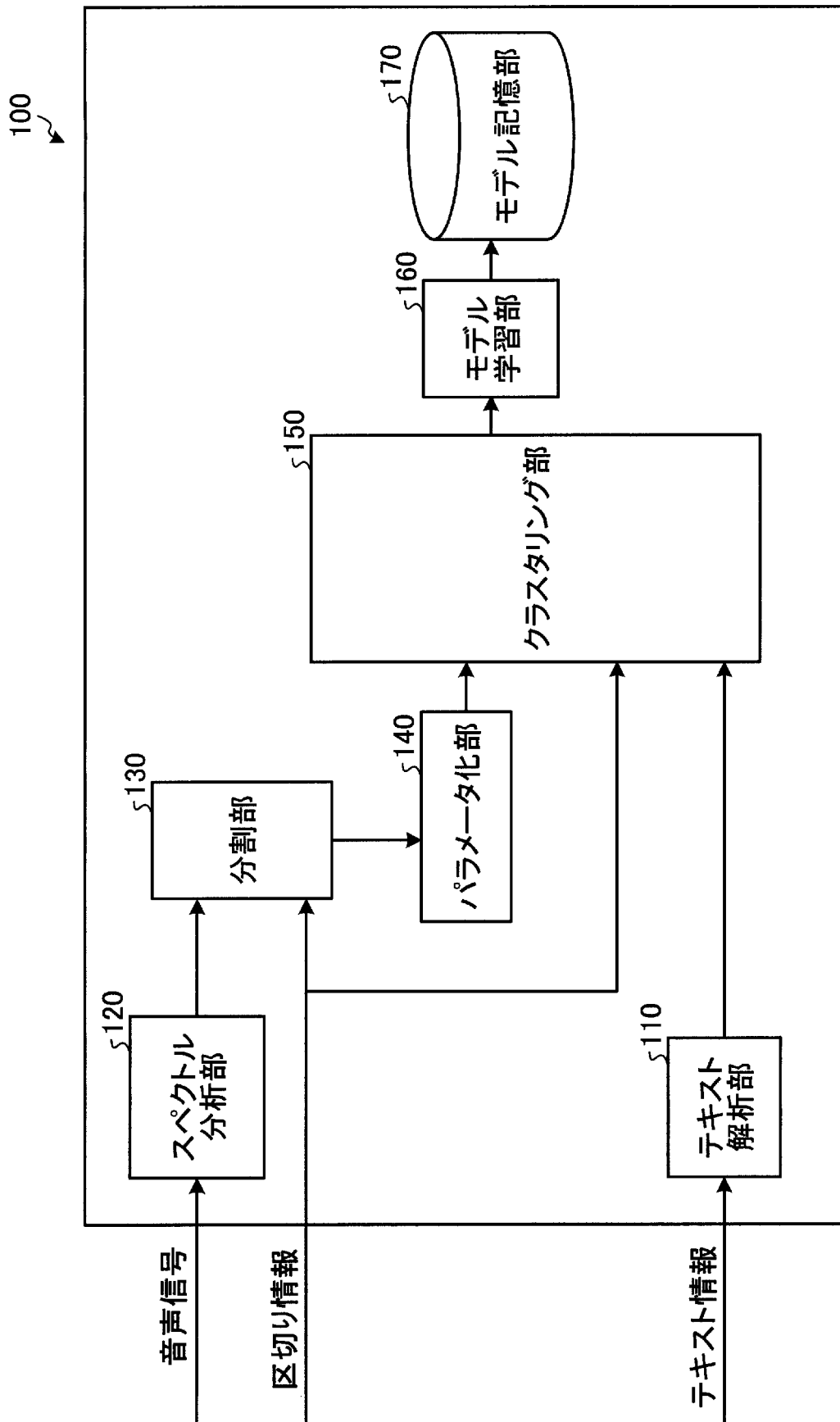
テキスト解析部が、音声合成の対象となるテキスト情報を取得し、前記テキスト情報をテキスト解析することにより、前記テキスト情報に含まれる言語の内容を示す言語情報を生成するテキスト解析ステップと、

選択部が、複数フレームを有する言語レベルを単位とする言語区間に含まれる、前記テキスト情報に対応する複数の音声信号それぞれの複数のスペクトルパラメータの特徴を示すスペクトルモデルであって、前記言語区間の前記言語情報により複数のクラスターにクラスタリングされたスペクトルモデルを記憶する記憶部から、前記音声合成の対象となるテキスト情報の前記言語区間の前記言語情報に基づいて、前記テキスト情報の前記言語区間が属する前記クラスターの前記スペクトルモデルを選択する選択ステップと、

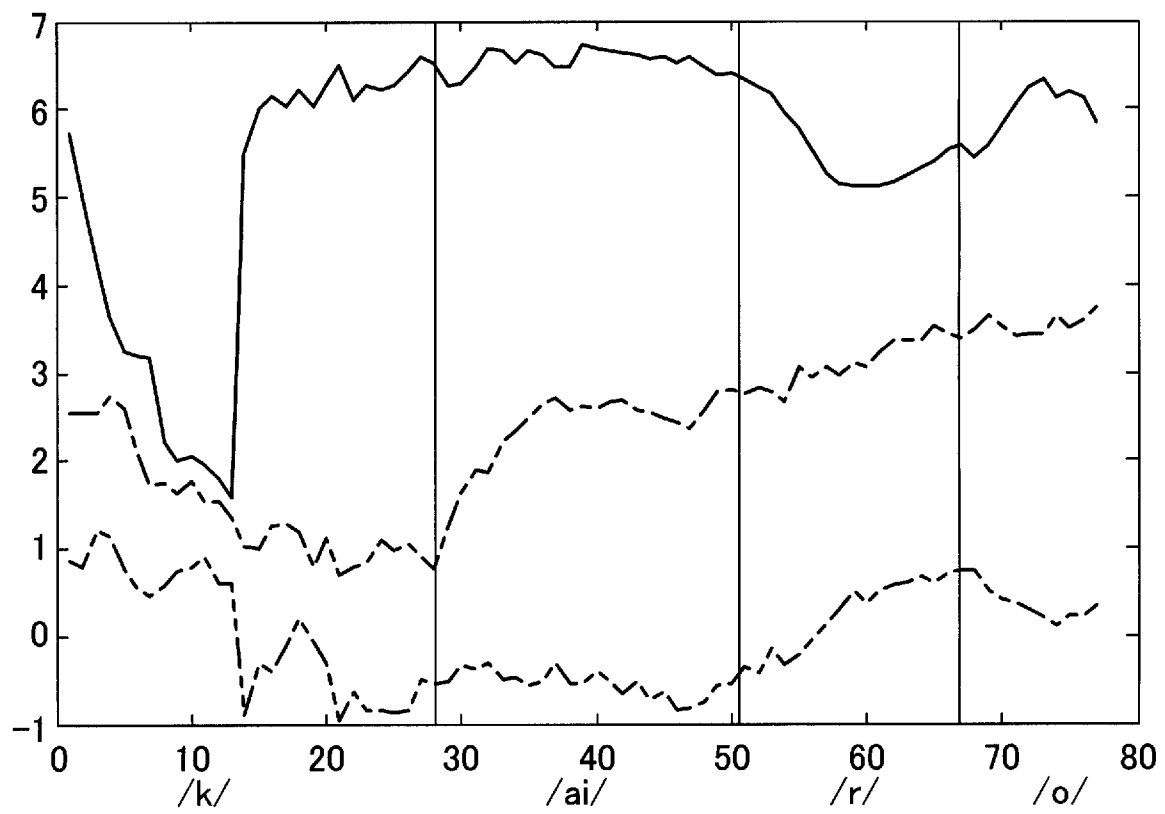
生成部が、前記選択ステップで選択された前記スペクトルモデルに基づいて、前記言語区間に対するスペクトルパラメータを生成し、前記スペクトルパラメータを逆変換することにより、特徴パラメータを得る生成ステップと

を有することを特徴とする音声合成方法。

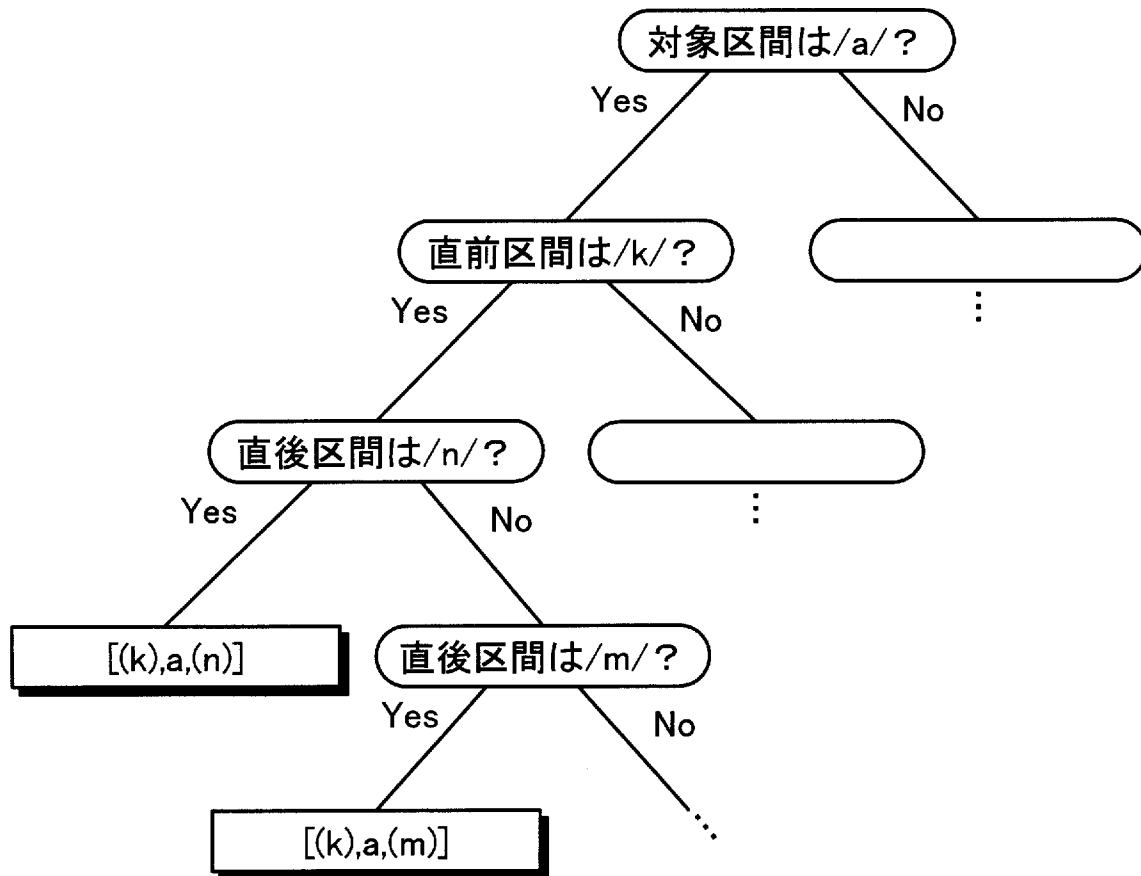
[図1]



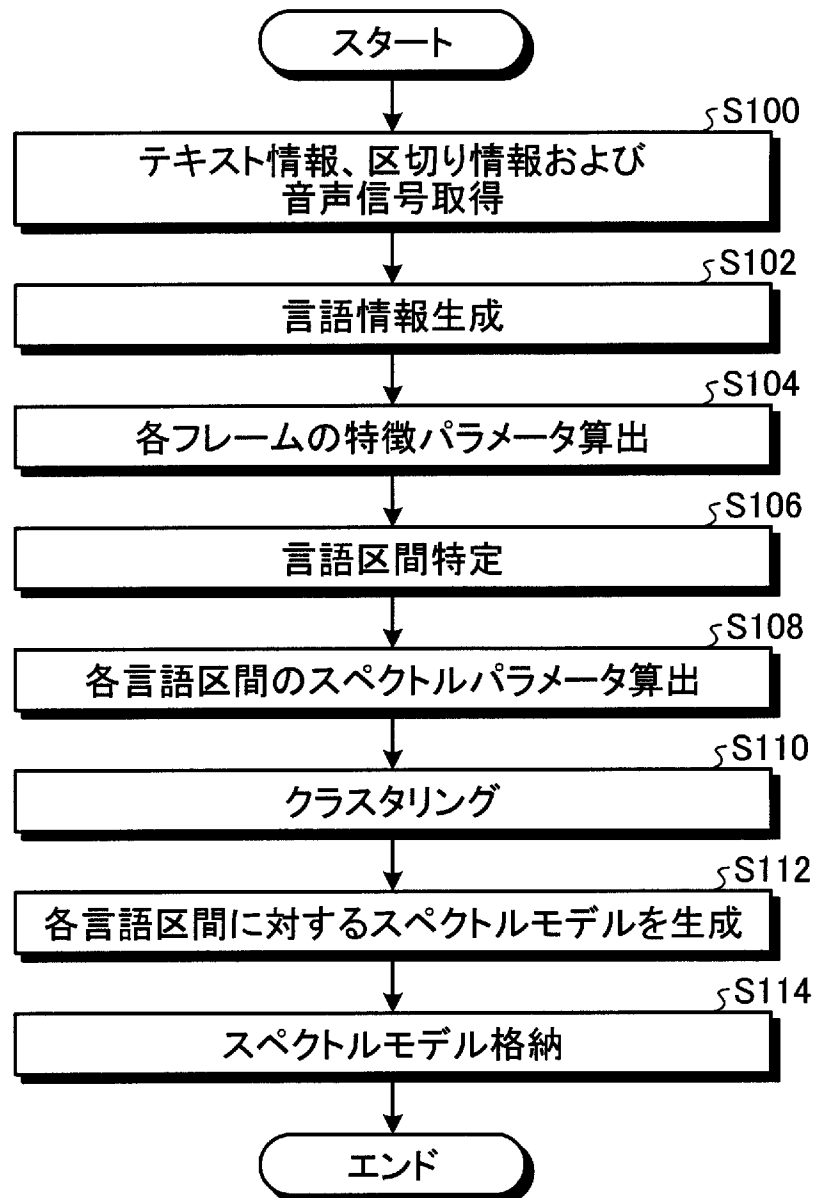
[図2]



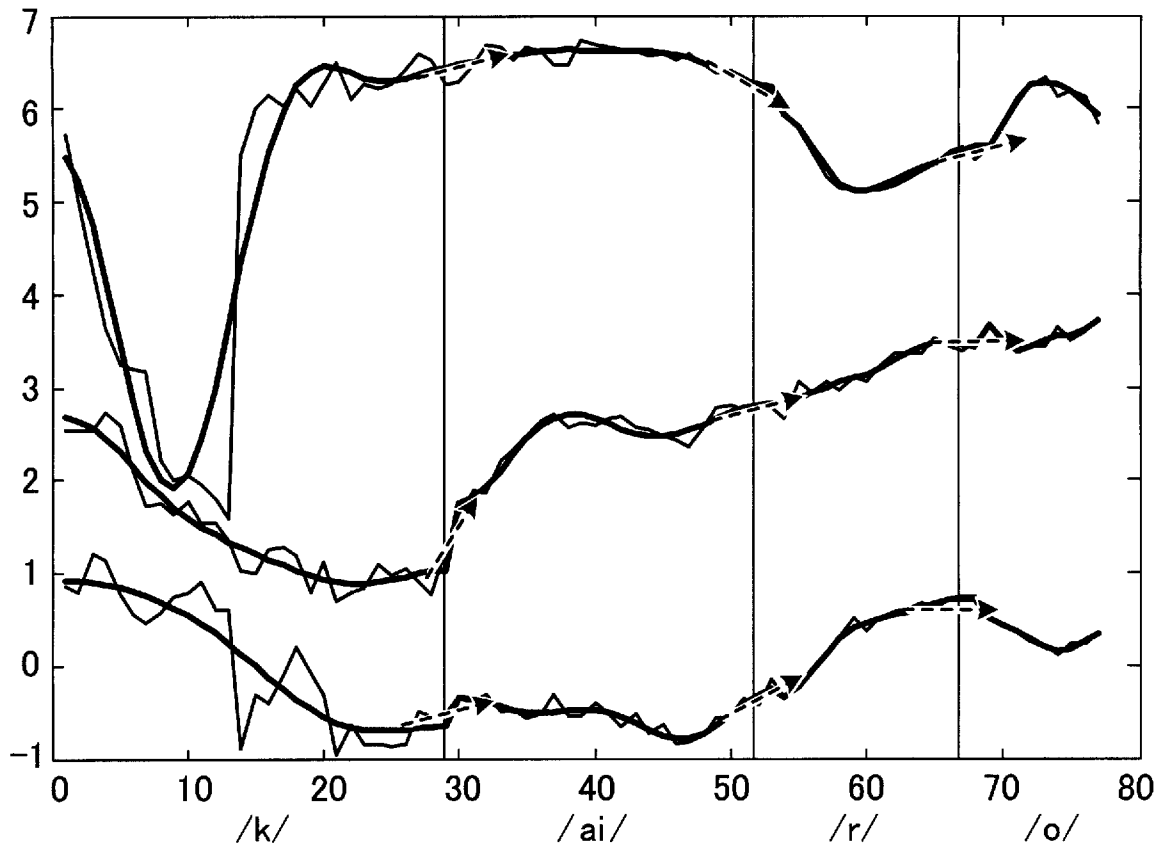
[図3]



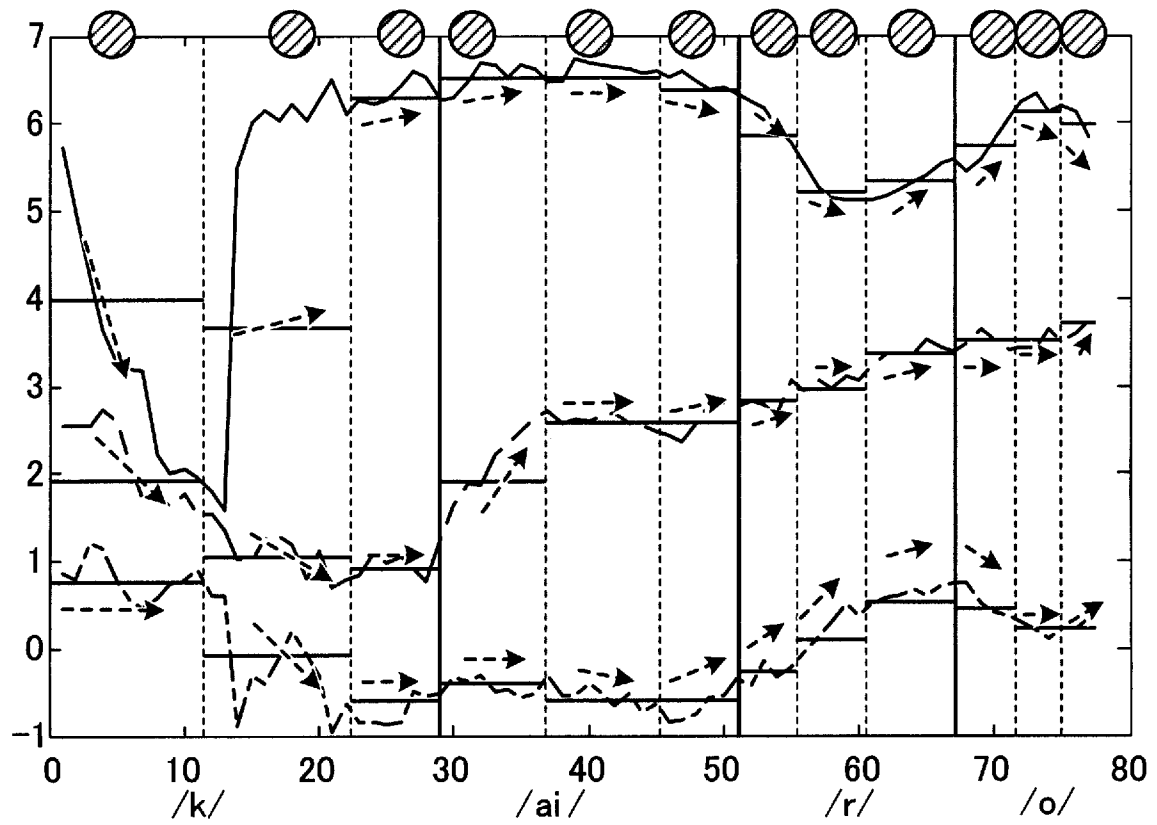
[図4]



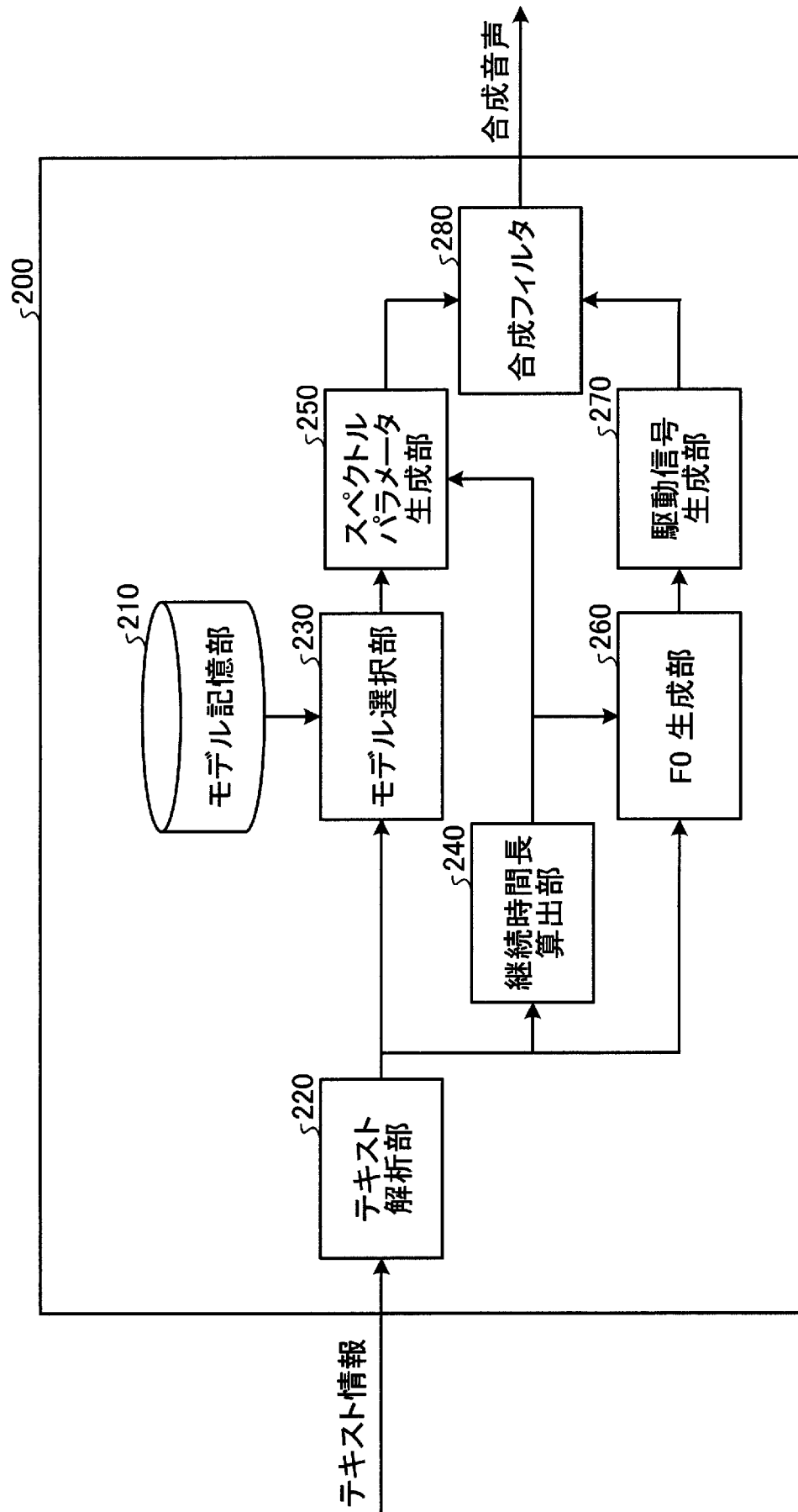
[図5]



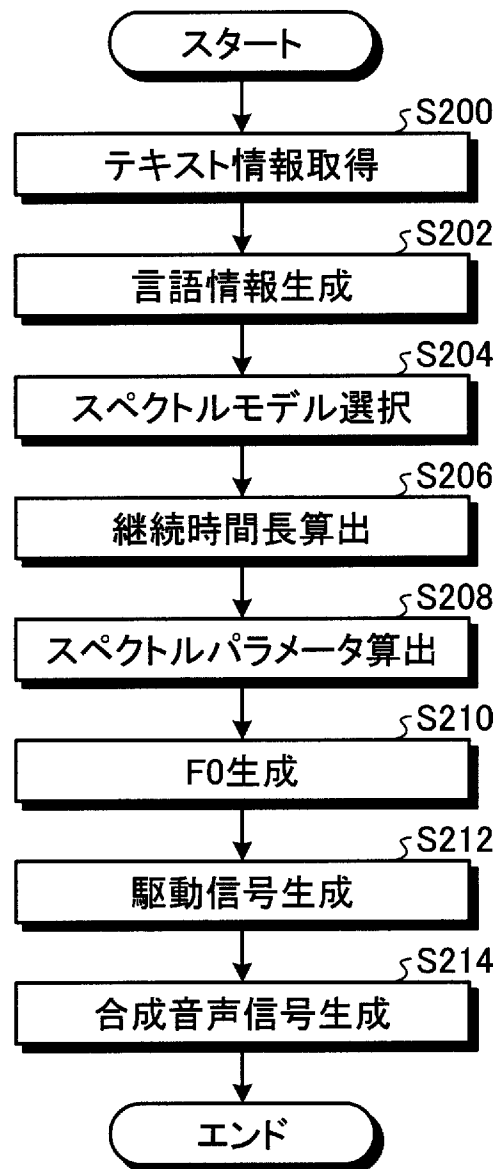
[図6]



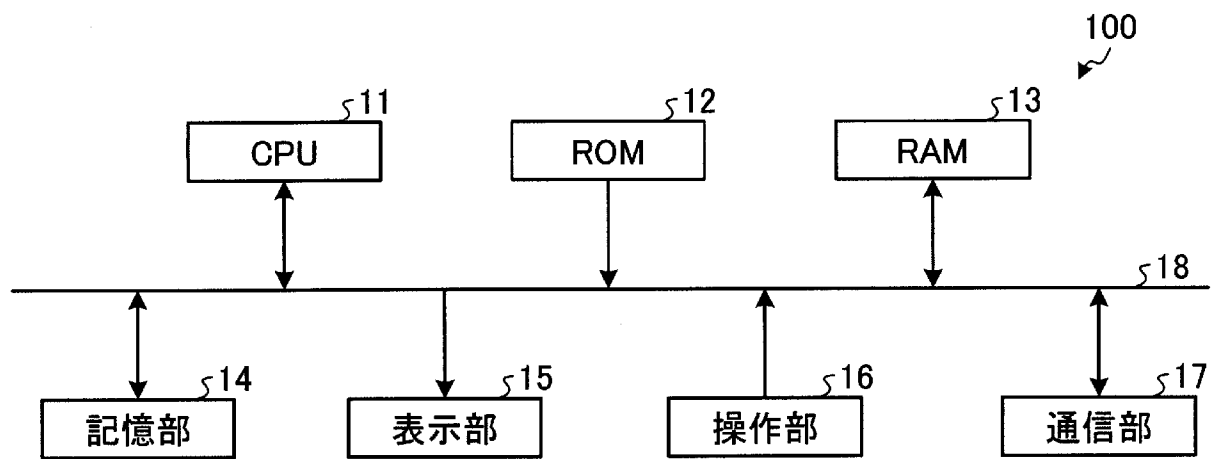
[図7]



[図8]



[図9]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2009/067408

A. CLASSIFICATION OF SUBJECT MATTER

G10L13/06 (2006.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G10L13/06

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho	1922-1996	Jitsuyo Shinan Toroku Koho	1996-2009
Kokai Jitsuyo Shinan Koho	1971-2009	Toroku Jitsuyo Shinan Koho	1994-2009

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

JSTPlus (JDreamII)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	JP 5-73100 A (Canon Inc.), 26 March 1993 (26.03.1993), entire text; all drawings & US 5633984 A	1, 4-10
Y	JP 8-69299 A (Sony Corp.), 12 March 1996 (12.03.1996), entire text; all drawings & US 5749065 A	1, 4-10
A	JP 9-258779 A (ATR Interpreting Telecommunications Research Laboratories), 03 October 1997 (03.10.1997), entire text; all drawings (Family: none)	1-10



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

16 December, 2009 (16.12.09)

Date of mailing of the international search report

28 December, 2009 (28.12.09)

Name and mailing address of the ISA/
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2009/067408

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2003-66983 A (Sharp Corp.), 05 March 2003 (05.03.2003), entire text; all drawings (Family: none)	1-10
A	JP 8-263095 A (NTT Data Communications Systems Corp.), 11 October 1996 (11.10.1996), entire text; all drawings (Family: none)	1-10
A	Jun'ichi YAMAGISHI et al., "Heikin Koe Model Kochiku ni Okeru Context Clustering to Washa Tekio Gakushu no Kento", IEICE Technical Report, 23 August 2002 (23.08.2002), vol.102, no.292, pages 5 to 10	1-10

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2009/067408

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

The inventions of claims 1 to 4, 7, and 9 relate to a sound model generation apparatus.

The inventions of claims 5, 6, 8, and 10 relate to a sound synthesis apparatus and are not considered to have such a configuration as to perform the specific operation for achieving sound synthesis using the sound model generated by the inventions of claims 1 to 4, 7, and 9.

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☒ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.

A. 発明の属する分野の分類（国際特許分類（I P C））

Int.Cl. G10L13/06(2006.01)i

B. 調査を行った分野

調査を行った最小限資料（国際特許分類（I P C））

Int.Cl. G10L13/06

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	1 9 2 2 - 1 9 9 6 年
日本国公開実用新案公報	1 9 7 1 - 2 0 0 9 年
日本国実用新案登録公報	1 9 9 6 - 2 0 0 9 年
日本国登録実用新案公報	1 9 9 4 - 2 0 0 9 年

国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）

JSTPlus(JDreamII)

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
Y	JP 5-73100 A（キヤノン株式会社）1993. 03. 26, 全文, 全図 & US 5633984 A	1, 4-10
Y	JP 8-69299 A（ソニー株式会社）1996. 03. 12, 全文, 全図 & US 5749065 A	1, 4-10

☒ C 欄の続きにも文献が列挙されている。☐ パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー

「A」特に関連のある文献ではなく、一般的技術水準を示すもの
「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）
「O」口頭による開示、使用、展示等に言及する文献
「P」国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献

「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの

「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの

「Y」特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの

「&」同一パテントファミリー文献

国際調査を完了した日

1 6 . 1 2 . 2 0 0 9

国際調査報告の発送日

2 8 . 1 2 . 2 0 0 9

国際調査機関の名称及びあて先

日本国特許庁（I S A / J P）

郵便番号 1 0 0 - 8 9 1 5

東京都千代田区霞が関三丁目4番3号

特許庁審査官（権限のある職員）

菊池 智紀

5 Z

3 3 5 2

電話番号 0 3 - 3 5 8 1 - 1 1 0 1 内線 3 5 4 1

C (続き) . 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
A	JP 9-258779 A (株式会社エイ・ティ・アール音声翻訳通信研究所) 1997. 10. 03, 全文, 全図 (ファミリーなし)	1-10
A	JP 2003-66983 A (シャープ株式会社) 2003. 03. 05, 全文, 全図 (ファミリーなし)	1-10
A	JP 8-263095 A (エヌ・ティ・ティ・データ通信株式会社) 1996. 10. 11, 全文, 全図 (ファミリーなし)	1-10
A	山岸順一 他, “平均声モデル構築におけるコンテキストクラスタリングと話者適 応学習の検討”, 電子情報通信学会技術研究報告, 2002. 08. 23, Vol. 102, No. 292, p. 5-10	1-10

第Ⅱ欄 請求の範囲の一部の調査ができないときの意見（第1ページの2の続き）

法第8条第3項（PCT17条(2)(a)）の規定により、この国際調査報告は次の理由により請求の範囲の一部について作成しなかった。

1. ☐ 請求項 _____ は、この国際調査機関が調査をすることを要しない対象に係るものである。
つまり、
2. ☐ 請求項 _____ は、有意義な国際調査をすることができる程度まで所定の要件を満たしていない国際出願の部分に係るものである。つまり、
3. ☐ 請求項 _____ は、従属請求の範囲であってPCT規則6.4(a)の第2文及び第3文の規定に従って記載されていない。

第Ⅲ欄 発明の単一性が欠如しているときの意見（第1ページの3の続き）

次に述べるようにこの国際出願に二以上の発明があるところの国際調査機関は認めた。

請求項1－4、7、9に係る発明は、音声モデル生成装置に関するものである。

請求項5、6、8、10に係る発明は、音声合成装置に関するものであり、請求項1－4、7、9に係る発明によって生成される音声モデルを用いて音声合成を行うための、特有の動作を行うような構成を有しているとは認められない。

1. ☐ 出願人が必要な追加調査手数料をすべて期間内に納付したので、この国際調査報告は、すべての調査可能な請求項について作成した。
2. ☒ 追加調査手数料を要求するまでもなく、すべての調査可能な請求項について調査することができたので、追加調査手数料の納付を求めなかった。
3. ☐ 出願人が必要な追加調査手数料を一部のみしか期間内に納付しなかったため、この国際調査報告は、手数料の納付のあった次の請求項のみについて作成した。
4. ☐ 出願人が必要な追加調査手数料を期間内に納付しなかったため、この国際調査報告は、請求の範囲の最初に記載されている発明に係る次の請求項について作成した。

追加調査手数料の異議の申立てに関する注意

- ☐ 追加調査手数料及び、該当する場合には、異議申立手数料の納付と共に、出願人から異議申立てがあった。
- ☐ 追加調査手数料の納付と共に出願人から異議申立てがあったが、異議申立手数料が納付命令書に示した期間内に支払われなかった。
- ☐ 追加調査手数料の納付はあったが、異議申立てはなかった。