(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
2006 年12 月21 日 (21.12.2006)

PCT

(10) 国際公開番号
WO 2006/134736 A1

(51) 国際特許分類:
G10L 13/08 (2006.01)

(21) 国際出願番号: PCT/JP2006/309288

(22) 国際出願日: 2006 年5 月9 日 (09.05.2006)

(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2005-176974 2005 年6 月16 日 (16.06.2005) JP

(71) 出願人 (米国を除く全ての指定国について): 松下電器産業株式会社 (MATSUSHITA ELECTRIC INDUSTRIAL CO., LTD.) [JP/JP]; 〒5718501 大阪府門真市大字門真１００６番地 Osaka (JP).

(72) 発明者; および

(75) 発明者/出願人 (米国についてのみ): 廣瀬 良文 (HIROSE, Yoshifumi). 釜井 孝浩 (KAMAI, Takahiro). 加藤 弓子 (KATO, Yumiko). 齋藤 夏樹 (SAITO, Natsuki).

(74) 代理人: 新居 広守 (NII, Hiromori); 〒5320011 大阪府大阪市淀川区西中島５丁目３番１０号 タナカ・イトーピア新大阪ビル６階 新居国際特許事務所内 Osaka (JP).

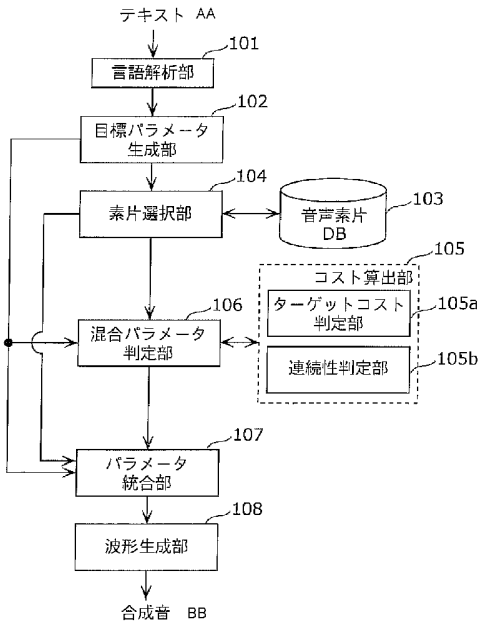
(81) 指定国 (表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, 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, HU, IE,

[続葉有]

(54) Title: SPEECH SYNTHESIZER, SPEECH SYNTHESIZING METHOD, AND PROGRAM

(54) 発明の名称: 音声合成装置、音声合成方法およびプログラム



AA... TEXT
101... LANGUAGE ANALYZING SECTION
102... TARGET PARAMETER GENERATING SECTION
104... FRAGMENT SELECTING SECTION
103... SPEECH FRAGMENT DB
106... MIX PARAMETER JUDGING SECTION
105... COST CALCULATING UNIT
105a... TARGET COST JUDGING SECTION
105b... CONTINUITY JUDGING SECTION
107... PARAMETER INTEGRATING SECTION
108... WAVEFORM CREATING SECTION
BB... SYNTHESIZED SOUND

(57) Abstract: A speech synthesizer for providing a synthesized sound of high and stable sound quality. The speech synthesizer comprises a target parameter generating section (102), a speech fragment DB (103), a fragment selecting section (104), a mix parameter judging section (106) for judging a combination of a target parameter and an optimal parameter of a speech fragment, a parameter integrating section (107) for integrating parameters, and a waveform creating section (108) for creating a synthesized sound. By combining a stable sound quality parameter which is generated by the target parameter generating section (102) and a speech fragment which is selected by the fragment selecting section (104), imparts an excellent sensation of real voice, and has a high sound quality for each parameter dimension, a high sound quality, stable synthesized sound is produced.

(57) 要約: 高音質で且つ安定した音質の合成音を提供することができる音声合成装置は、目標パラメータ生成部（１０２）と、音声素片ＤＢ（１０３）と、素片選択部（１０４）と、目標パラメータと音声素片の最適なパラメータの組み合わせを判定する混合パラメータ判定部（１０６）と、パラメータを統合するパラメータ統合部（１０７）と、合成音を生成する波形生成部（１０８）を備え、目標パラメータ生成部（１０２）により生成される音質の安定したパラメータと、前記素片選択部（１０４）により選択される肉声感が高く音質の高い音声素片とをパラメータ次元毎に組み合わせることにより、高音質かつ安定した合成音を生成する。

WO 2006/134736 A1

IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

添付公開書類:
— 国際調査報告書

２文字コード及び他の略語については、定期発行される各*PCT*ガゼットの巻頭に掲載されている「コードと略語のガイダンスノート」を参照。

明 細 書

音声合成装置、音声合成方法およびプログラム

技術分野

[0001] 本発明は、高音質で、かつ安定した音質の合成音を提供する音声合成装置に関するものである。

背景技術

[0002] 従来の肉声感の高い音声合成装置としては、大規模な素片DBから波形を選択して接続する波形接続方式を用いるものがあった(例えば、特許文献1参照)。図1は、波形接続型音声合成装置の典型的な構成図である。

[0003] 波形接続型音声合成装置は、入力されたテキストを合成音声に変換する装置であり、言語解析部101と、韻律生成部201と、音声素片DB(データベース)202と、素片選択部104と、波形接続部203とを備えている。

[0004] 言語解析部101は、入力されたテキストを言語的に解析し、発音記号およびアクセント情報を出力する。韻律生成部201は、言語解析部101より出力された発音記号およびアクセント情報に基づいて、発音記号毎に基本周波数、継続時間長、パワーなどの韻律情報を生成する。音声素片DB202は、予め収録された音声波形を保持する。素片選択部104は、韻律生成部201により生成された韻律情報に基づいて、音声素片DB202より最適な音声素片を選択する処理部である。波形接続部203は、素片選択部104により選択された音声素片を接続し、合成音声を生成する。

[0005] また、安定した音質の音声を提供する音声合成装置としては、統計モデルを学習することにより合成パラメータを生成し、音声を合成する装置も知られている(例えば、特許文献2参照)。図2は、統計モデルによる音声合成方式の一つであるHMM(隠れマルコフモデル)音声合成方式を用いた音声合成装置の構成図である。

[0006] 音声合成装置は、学習部100および音声合成部200から構成される。学習部100は、音声DB202、励振源スペクトルパラメータ抽出部401、スペクトルパラメータ抽出部402およびHMMの学習部403を備えている。また、音声合成部200は、コンテキスト依存HMMファイル301、言語解析部101、HMMからのパラメータ生成部404、

励振源生成部405および合成フィルタ303を備えている。

[0007] 学習部100は、音声DB202に格納された音声情報よりコンテキスト依存HMMファイル301を学習させる機能をもつ。音声DB202には、あらかじめサンプルとして用意された多数の音声情報が格納されている。音声情報は、図示の例のように、音声信号に波形の各音素等の部分を識別するラベル(arayuruやnuuyooku)を付加したものである。励振源スペクトルパラメータ抽出部401およびスペクトルパラメータ抽出部402は、それぞれ音声DB202から取り出した音声信号ごとに、励振源パラメータ列およびスペクトルパラメータ列を抽出する。HMMの学習部403は、抽出された励振源パラメータ列およびスペクトルパラメータ列について、音声DB202から音声信号とともに取り出したラベルおよび時間情報を用いて、HMMの学習処理を行なう。学習されたHMMは、コンテキスト依存HMMファイル301に格納される。励振源モデルのパラメータは、多空間分布HMMを用いて学習を行う。多空間分布HMMは、パラメータベクトルの次元が、毎回、異なることを許すように拡張されたHMMであり、有声／無声フラグを含んだピッチは、このような次元が変化するパラメータ列の例である。つまり、有声時には1次元、無声時には0次元のパラメータベクトルとなる。学習部100では、この多空間分布HMMによる学習を行っている。ラベル情報とは、具体的には、例えば、以下のようなものを指し、各HMMは、これらを属性名(コンテキスト)として持つ。

・{先行、当該、後続}音素
・当該音素のアクセント句内でのモーラ位置
・{先行、当該, 後続}の品詞, 活用形, 活用型
・{先行, 当該、後続}アクセント句のモーラ長, アクセント型
・当該アクセント句の位置, 前後のポーズの有無
・{先行, 当該, 後続}呼気段落のモーラ長
・当該呼気段落の位置
・文のモーラ長

このようなHMMは、コンテキスト依存HMMと呼ばれる。

[0008] 音声合成部200は、任意の電子的なテキストから読み上げ形式の音声信号列を生

成する機能をもつ。言語解析部101は、入力されたテキストを解析して、音素の配列であるラベル情報に変換する。HMMからのパラメータ生成部404は、言語解析部101より出力されるラベル情報に基づいてコンテキスト依存HMMファイル301を検索する。そして、得られたコンテキスト依存HMMを接続し、文HMMを構成する。励振源生成部405は、得られた文HMMから、さらにパラメータ生成アルゴリズムにより、励振源パラメータを生成する。また、HMMからのパラメータ生成部404は、スペクトルパラメータの列を生成する。さらに、合成フィルタ303が、合成音を生成する。

[0009] また、実音声波形と、パラメータとを組み合わせる方法としては、例えば特許文献3の方法がある。図3は、特許文献3の音声合成装置の構成を示す図である。

[0010] 特許文献3の音声合成装置には音韻記号解析部1が設けられ、その出力は制御部2に接続されている。また、音声合成装置には個人情報DB10が設けられ、制御部2と互いに接続されている。さらに、音声合成装置には自然音声素片チャンネル12と合成音声素片チャンネル11とが設けられている。自然音声素片チャンネル12の内部には音声素片DB6と音声素片読み出し部5とが設けられている。合成音声素片チャンネル11の内部にも同様に音声素片DB4と音声素片読み出し部3とが設けられている。音声素片読み出し部5は音声素片DB6と互いに接続されている。音声素片読み出し部3は音声素片DB4と互いに接続されている。音声素片読み出し部3と音声素片読み出し部5との出力は混合部7の二つの入力に接続されており、混合部7の出力は振幅制御部8に入力されている。振幅制御部8の出力は出力部9に入力されている。

[0011] 制御部2からは各種の制御情報が出力される。制御情報には自然音声素片インデックス、合成音声素片ンデックス、混合制御情報および振幅制御情報が含まれる。まず、自然音声素片インデックスは自然音声素片チャンネル12の音声素片読み出し部5に入力されている。合成音声素片インデックスは合成音声素片チャンネル11の音声素片読み出し部3に入力されている。混合制御情報は混合部7に入力されている。そして、振幅制御情報は振幅制御部8に入力されている。

[0012] この方法では、予め作成しておいたパラメータによる合成素片と、収録された合成素片とを混合する方法として、自然音声素片と合成音声素片の双方をCV単位(日本

語の1音節に対応する一対の子音と母音の組み合わせの単位)などで時間的に比率を変更しながら混合する。よって、自然音声素片を用いた場合と比較して記憶量を削減でき、かつ、少ない計算量で、合成音を得ることができる。

特許文献1:特開平10－247097号公報(段落0007、図1)

特許文献2:特開2002－268660号公報(段落0008－0011、図1)

特許文献3:特開平9－62295号公報(段落0030－0031、図1)

発明の開示

発明が解決しようとする課題

[0013] しかしながら、前記従来の波形接続型音声合成装置(特許文献1)の構成では、音声素片DB202に予め保持されている音声素片だけしか音声合成に利用することが出来ない。つまり、韻律生成部201により生成された韻律に類似した音声素片がない場合には、韻律生成部201により生成された韻律とは、大きく異なる音声素片を選択せざるを得ない。したがって、局所的に音質が劣化するという課題を有している。また、音声素片DB202が十分に大きく構築できない場合は、上記課題が顕著に生じるという課題を有している。

[0014] 一方、前記従来の統計モデルによる音声合成装置(特許文献2)の構成では、予め収録された音声DB202により統計的に学習されたHMMモデル(隠れマルコフモデル)を用いることにより、言語解析部101により出力される発音記号およびアクセント情報のコンテキストラベルに基づいて、統計的に合成パラメータを生成する。そのため、全ての音韻において安定した音質の合成音を得ることが可能である。しかし、一方で、HMMモデルによる統計的な学習を用いていることにより、個々の音声波形が保有する微細な特徴(韻律の微細な変動で合成音声の自然さに影響を及ぼすマイクロプロソディなど)が統計処理によって失われるために合成音声の肉声感は低下し、鈍った音声になるという課題を有している。

[0015] また、前記従来のパラメータ統合方法では、合成音声素片と自然音声素片の混合は、CV間の過渡期に時間的に用いていた為、全時間にわたる均一な品質を得ることが困難であり、時間的に音声の質が変化するという課題が存在する。

[0016] 本発明は、前記従来の課題を解決するもので、高音質で且つ安定した音質の合成

音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

[0017] 本発明に係る音声合成装置は、少なくとも発音記号を含む情報から、音声を合成することが可能なパラメータ群である目標パラメータを素片単位で生成する目標パラメータ生成部と、予め録音された音声を、前記目標パラメータと同じ形式のパラメータ群からなる音声素片として素片単位で記憶している音声素片データベースと、前記目標パラメータに対応する音声素片を前記音声素片データベースより選択する素片選択部と、音声素片ごとに、前記目標パラメータのパラメータ群および前記音声素片のパラメータ群を統合してパラメータ群を合成するパラメータ群合成部と、合成された前記パラメータ群に基づいて、合成音波形を生成する波形生成部とを備えることを特徴とする。例えば、前記コスト算出部は、前記素片選択部により選択された音声素片の部分集合と、当該音声素片の部分集合に対応する前記目標パラメータの部分集合との非類似性を示すコストを算出するターゲットコスト判定部を有していてもよい。

[0018] 本構成によって、目標パラメータ生成部により生成される音質の安定したパラメータと、前記素片選択部により選択される肉声感が高く音質の高い音声素片とを組み合わせることにより、高音質かつ安定した音質の合成音を生成することができる。

[0019] また、前記パラメータ群合成部は、前記目標パラメータ生成部により生成された目標パラメータを、少なくとも1つ以上の部分集合に分割することによって得られるパラメータパターンを少なくとも1つ以上生成する目標パラメータパターン生成部と、前記目標パラメータパターン生成部により生成された前記目標パラメータの部分集合ごとに、当該部分集合に対応する音声素片を前記音声素片データベースより選択する素片選択部と、前記素片選択部により選択された音声素片の部分集合と当該音声素片の部分集合に対応する前記目標パラメータの部分集合とに基づいて、当該音声素片の部分集合を選択することによるコストを算出するコスト算出部と、前記コスト算出部によるコスト値に基づいて、前記目標パラメータの部分集合の最適な組み合わせを、素片ごとに判定する組み合わせ判定部と、前記組み合わせ判定部により判定された組み合わせに基づいて、前記素片選択部により選択された前記音声素片の部分集合を統合することによりパラメータ群を合成するパラメータ統合部とを有してい

てもよい。

[0020] 本構成によって、前記目標パラメータパターン生成部により生成される複数のパラメータの部分集合に基づいて、前記素片選択部により選択される肉声感が高く音質の高い音声素片のパラメータの部分集合を組み合わせ判定部により適切に組み合わせている。このため、高音質かつ安定した合成音を生成することができる。

発明の効果

[0021] 本発明の音声合成装置によれば、実音声に基づく音声素片データベースから選択した音声素片のパラメータと、統計モデルに基づく安定した音質のパラメータとを適宜混合することにより、安定でかつ高音質の合成音を得ることができる。

図面の簡単な説明

[0022] [図1]図1は、従来の波形接続型音声合成装置の構成図である。

[図2]図2は、従来の統計モデルに基づく音声合成装置の構成図である。

[図3]図3は、従来のパラメータ統合方法の構成図である。

[図4]図4は、本発明の実施の形態1における音声合成装置の構成図である。

[図5]図5は、音声素片の説明図である。

[図6]図6は、本発明の実施の形態1のフローチャートである。

[図7]図7は、パラメータ混合結果の説明図である。

[図8]図8は、混合パラメータ判定部のフローチャートである。

[図9]図9は、組み合わせベクトル候補生成の説明図である。

[図10]図10は、ビタビアルゴリズムの説明図である。

[図11]図11は、混合ベクトルをスカラー値にした場合のパラメータ混合結果を示す図である。

[図12]図12は、声質変換を行う場合の説明図である。

[図13]図13は、本発明の実施の形態2における音声合成装置の構成図である。

[図14]図14は、本発明の実施の形態2のフローチャートである。

[図15]図15は、目標パラメータパターン生成部の説明図である。

[図16]図16は、組み合わせベクトル判定部のフローチャートである。

[図17A]図17Aは、選択ベクトル候補生成の説明図である。

[図17B]図17Bは、選択ベクトル候補生成の説明図である。

[図18]図18は、組み合わせ結果の説明図である。

[図19]図19は、コンピュータの構成の一例を示す図である。

符号の説明

[0023] 1 音韻記号列解析部

2 制御部

3 音声素片読み出し部

4 音声素片DB

5 音声素片読み出し部

6 音声素片DB

7 混合部

8 振幅制御部

9 出力部

10 個人情報DB

11 合成音声素片チャンネル

12 自然音清素片チャンネル

41 目標パラメータを使用する領域

42 実音声パラメータを使用する領域

43 実音声パラメータを使用する領域

44 実音声パラメータを使用する領域

45 目標パラメータを使用する領域

100 学習部

200 音声合成部

101 言語解析部

102 目標パラメータ生成部

103 音声素片DB

104 素片選択部

105 コスト算出部

105a ターゲットコスト判定部

105b 連続性コスト判定部

106 混合パラメータ判定部

107 パラメータ統合部

108 波形生成部

201 韻律生成部

202 音声素片DB

203 波形接続部

301 コンテキスト依存HMMファイル

302 文章HMM作成部

303 合成フィルタ

401 励振源スペクトルパラメータ抽出部

402 スペクトルパラメータ抽出部

403 HMMの学習部

404 HMMからのパラメータ生成部

405 励振源生成部

601 実音声パラメータを使用する素片の領域

602 目標パラメータを使用する素片の領域

603 実音声パラメータを使用する素片の領域

604 目標パラメータを使用する素片の領域

801 目標パラメータパターン生成部

802 組み合わせ判定部

1101 標準音声DB

1102 感情音声DB

1501 パターンA1により選択された素片

1502 パターンC2により選択された素片

発明を実施するための最良の形態

[0024] 以下本発明の実施の形態について、図面を参照しながら説明する。

[0025] （実施の形態1）

図4は、本発明の実施の形態1における音声合成装置の構成図である。

[0026] 本実施の形態の音声合成装置は、高音質と音質の安定性とを両立させた音声を合成する装置であって、言語解析部101と、目標パラメータ生成部102と、音声素片DB103と、素片選択部104と、コスト算出部105と、混合パラメータ判定部106と、パラメータ統合部107と、波形生成部108とを備えている。コスト算出部105は、ターゲットコスト判定部105aと、連続性判定部105bとを備えている。

[0027] 言語解析部101は、入力されたテキストを解析し、発音記号やアクセント情報を出力する。例えば、「今日の天気は」というテキストが入力された場合、「kyo'－no／te'Nkiwa」といったような発音記号、およびアクセント情報を出力する。ここで、「'」はアクセント位置を示し、「／」はアクセント句境界を示す。

[0028] 目標パラメータ生成部102は、言語解析部101により出力された発音記号やアクセント情報に基づいて、音声を合成するために必要なパラメータ群を生成する。パラメータ群を生成する方法は特に限定するものではない。例えば、特許文献2に示されているようにHMM（隠れマルコフモデル）を用いることにより、安定した音質のパラメータを生成することが可能である。

[0029] 具体的には特許文献2に記載の方法を用いればよい。なおパラメータの生成方法はこれに限るものではない。

[0030] 音声素片DB103は、予め収録した音声（自然音声）を分析し、再合成可能なパラメータ群として保持するデータベースである。また、保持する単位を素片と呼ぶ。素片の単位は特に限定するものではなく、音素、音節、モーラ、アクセント句などを用いればよい。本発明の実施の形態では、素片の単位として音素を用いて説明する。また、パラメータの種類は特に限定するものではないが、例えば、パワー、継続時間長、基本周波数といった音源情報と、ケプストラムなどの声道情報をパラメータ化し保持すればよい。1つの音声素片は、図5に示すように複数フレームのk次元のパラメータで表現される。図5では、素片P_iは、mフレームにより構成されており、各フレームはk個のパラメータにより構成される。このようにして構成されるパラメータにより音声を再合成することが可能となる。例えば、図中、$P_{i1}=(p_{11}, p_{21}, p_{31}, \cdots, p_{m1})$と示されている

のは、素片P_iにおける1番目のパラメータのmフレームにわたる時間変化を示している。

[0031] 素片選択部104は、目標パラメータ生成部102により生成された、目標パラメータに基づいて、音声素片DB103から、音声素片系列を選択する選択部である。

[0032] ターゲットコスト判定部105aは目標パラメータ生成部102により生成された目標パラメータと、素片選択部104により選択された音声素片との類似度に基づくコストを、素片単位ごとに算出する。

[0033] 連続性判定部105bは、素片選択部104により選択された音声素片のパラメータの一部を、目標パラメータ生成部102により生成された目標パラメータで置き換える。そして、音声素片を接続した場合に起こる歪み、つまりパラメータの連続性を算出する。

[0034] 混合パラメータ判定部106は、ターゲットコスト判定部105aと連続性判定部105bとにより算出されるコスト値に基づいて、音声合成時に使用するパラメータとして、音声素片DB103より選択したパラメータを用いるか、目標パラメータ生成部102により生成されたパラメータを用いるかを示す選択ベクトルを素片単位毎に決定する。混合パラメータ判定部106の動作は後で詳述する。

[0035] パラメータ統合部107は混合パラメータ判定部106により決定された選択ベクトルに基づいて、音声素片DB103より選択されたパラメータと目標パラメータ生成部102により生成されたパラメータとを統合する。

[0036] 波形生成部108は、パラメータ統合部107により生成された合成パラメータに基づいて合成音を合成する。

[0037] 上記のように構成した音声合成装置の動作について、次に詳述する。

[0038] 図6は、音声合成装置の動作の流れを示すフローチャートである。言語解析部101は、入力されたテキストを言語的に解析し、発音記号およびアクセント記号を生成する(ステップS101)。目標パラメータ生成部102は、発音記号およびアクセント記号に基づいて、上述のHMM音声合成法により、再合成可能なパラメータ系列$T=t_1, t_2, \cdots, t_n$を生成する(nは素片数)(ステップS102)。以後、この目標パラメータ生成部102により生成されたパラメータ系列を目標パラメータと呼ぶ。

[0039] 素片選択部104は、生成された目標パラメータに基づいて、音声素片DB103から目標パラメータに最も近い音声素片系列U＝$u_1, u_2, \cdots, u_n$を選択する(ステップS103)。以降、選択された音声素片系列を実音声パラメータと呼ぶ。選択の方法は特に限定するものではないが、例えば、特許文献1に記載の方法により選択することが可能である。

[0040] 混合パラメータ判定部106は、目標パラメータと実音声パラメータとを入力とし、パラメータの次元毎にどちらのパラメータを使用するかを示す選択ベクトル系列Cを決定する(ステップS104)。選択ベクトル系列Cは、式1に示すように素片ごとの選択ベクトルC_iからなる。選択ベクトルC_iは、i番目の素片について、パラメータ次元毎に目標パラメータと実音声パラメータのどちらを使用するかを2値で示している。例えば、c_{ij}が0の場合には、i番目の素片のj番目のパラメータについては、目標パラメータを使用する。また、c_{ij}が1の場合には、i番目の素片のj番目のパラメータについては、音声素片DB103より選択された実音声パラメータを使用することを示している。

[0041] 図7は、選択ベクトル系列Cによって、目標パラメータと、実音声パラメータとを切り分けた例である。図7には、実音声パラメータを使用する領域42、43および44と、目標パラメータを使用する領域41および45とが示されている。例えば、1番目の素片P_1からP_{k1}に着目すると、1番目のパラメータについては、目標パラメータを使用し、2番目からk番目のパラメータについては、実音声パラメータを使用することが示されている。

[0042] この選択ベクトル系列Cを適切に決定することにより、目標パラメータによる安定した音質と、実音声パラメータによる肉声感の高い高音質とを両立する高音質且つ安定した合成音を生成することが可能になる。

[0043] [数1]

$$C = C_1, C_2, \cdots, C_n$$

但し

$$C_i = c_{i1}, c_{i2}, \cdots, c_{ik}$$

$$c_{ij} = \begin{cases} 0 & \text{目標パラメータを使用する場合} \\ 1 & \text{実音声パラメータを使用する場合} \end{cases} \qquad \text{(式1)}$$

[0044] 次に選択ベクトル系列Cの決定方法(図6のステップS104)について説明する。混合パラメータ判定部106は、高音質で且つ安定し合成音を生成する為に、実音声パラメータが目標パラメータに類似している場合は、実音声パラメータを使用し、類似していない場合は目標パラメータを使用する。また、この時、目標パラメータとの類似度だけではなく、前後の素片との連続性を考慮する。これにより、パラメータの入替えによる不連続を軽減することが可能である。この条件を満たす選択ベクトル系列Cは、ビタビアルゴリズムを用いて探索する。

[0045] 探索アルゴリズムを図8に示すフローチャートを用いて説明する。素片i=1, …, nに対して順次ステップS201からステップS205までの処理が繰り返される。

[0046] 混合パラメータ判定部106は、対象となる素片に対して、選択ベクトルC_iの候補h_iとして、p個の候補$h_{i,1}$, $h_{i,2}$, …, $h_{i,p}$を生成する(ステップS201)。生成する方法は特に限定するものではない。例えば、生成方法として、k次元のそれぞれのパラメータに対しての全ての組み合わせを生成しても構わない。また、より効率的に候補の生成を行うために、図9に示すように、1つ前の選択ベクトルC_{i-1}との差分が所定の閾値以下になるような組み合わせのみを生成するようにしても構わない。また、最初の素片(i=1)に関しては、例えば、全て目標パラメータを使用するような候補を生成してもよいし($C_1=(0, 0, \cdots, 0)$)、逆に全て実音声パラメータを使用するような候補を生成するようにしてもよい($C_1=(1, 1, \cdots, 1)$)。

[0047] ターゲットコスト判定部105aは、選択ベクトルC_iのp個の候補$h_{i,1}$, $h_{i,2}$, …, $h_{i,p}$の各々について、目標パラメータ生成部102により生成された目標パラメータt_iと、素片選択部104により選択された音声素片u_iとの類似度に基づくコストを、式2により計算する(ステップS202)。

[0048] [数2]

$$TargetCost(h_{i,j}) = \omega_1 \times Tc(h_{i,j} \bullet u_i, h_{i,j} \bullet t_i) + \omega_2 \times Tc((1-h_{i,j}) \bullet u_i, (1-h_{i,j}) \bullet t_i)$$

ただし、 j = 1 ～ p … (式2)

[0049] ここで、ω_1, ω_2は、重みであり、$\omega_1 > \omega_2$とする。重みの決定方法は特に限定するものではないが、経験に基づき決定することが可能である。また、$h_{i,j} \cdot u_i$は、は、ベクトル$h_{i,j}$とベクトルu_iの内積であり、実音声パラメータu_iのうち、選択ベクトル候補$h_{i,j}$によって

採用される部分パラメータ集合を示す。一方、$(1-h_{i,j})\cdot u_i$は、実音声パラメータu_iのうち、選択ベクトル候補$h_{i,j}$によって採用されなかった部分パラメータ集合を示す。目標パラメータt_iについても同様である。関数Tcは、パラメータ間の類似度に基づくコスト値を算出する。算出方法は特に限定するものではないが、例えば、各パラメータ次元間の差分の重み付け加算により算出することが可能である。例えば、類似度が大きくなるほどコスト値が小さくなるように関数Tcが定められている。

[0050] 繰り返すと、式2の1項目の関数Tcの値は、選択候補ベクトル$h_{i,j}$によって採用された、実音声パラメータu_iの部分パラメータ集合および目標パラメータt_iの部分パラメータ集合同士の類似度に基づくコスト値を示す。式2の2項目の関数Tcの値は、選択候補ベクトル$h_{i,j}$によって採用されなかった実音声パラメータu_iの部分パラメータ集合、および目標パラメータt_iの部分パラメータ集合同士の類似度に基づくコスト値を示している。式2はこれら2つのコスト値の重み付け和を示したものである。

[0051] 連続性判定部105bは、選択ベクトル候補$h_{i,j}$それぞれについて、1つ前の選択ベクトル候補との連続性に基づくコストを式3を用いて評価する(ステップS203)。

[0052] [数3]

$$ContCost(h_{i,j},h_{i-1,r})=Cc(h_{i,j}\bullet u_i+(1-h_{i,j})\bullet t_i,h_{i-1,r}\bullet u_{i-1}+(1-h_{i-1,r})\bullet t_{i-1})$$

(式3)

[0053] ここで、$h_{i,j}\cdot u_i+(1-h_{i,j})\cdot u_i$は、選択ベクトル候補$h_{i,j}$によって規定される目標パラメータ部分集合と、実音声パラメータ部分集合の組み合わせによって構成される素片iを形成するパラメータであり、$h_{i-1,r}\cdot u_{i-1}+(1-h_{i-1,r})\cdot u_{i-1}$は、1つ前の素片i−1に対する選択ベクトル候補$h_{i-1,r}$により規定される素片i−1を形成するパラメータである。

[0054] 関数Ccは、2つの素片パラメータの連続性に基づくコストを評価する関数である。すなわち、2つの素片パラメータの連続性がよい場合には、値が小さくなる関数である。算出方法は特に限定するものではないが、例えば、素片i−1の最終フレームと素片iの先頭フレームにおける各パラメータ次元の差分値の重み付け和により計算すればよい。

[0055] 混合パラメータ判定部106は、図10に示すように、式4に基づいて選択ベクトル候補$h_{i,j}$に対するコスト($C(h_{i,j})$)を算定し、同時に素片i−1に対する選択ベクトル候補h_i

$_{-1,r}$のうちどの選択ベクトル候補と接続すべきかを示す接続元（B（$h_{i,j}$））を決定する（ステップS204）。なお、図10では、接続元として$h_{i-1,3}$が選択されている）。

[0056] [数4]

$$C(h_{i,j}) = \mathrm{TargetCost}(h_{i,j}) + \underset{p}{\mathrm{Min}}\left[\mathrm{ContCost}(h_{i,j}, h_{i-1,p}) + C(h_{i-1,p})\right]$$
$$B(h_{i,j}) = \underset{p}{\arg\min}\left[\mathrm{ContCost}(h_{i,j}, h_{i-1,p}) + C(h_{i-1,p})\right] \quad \text{（式４）}$$

[0057] ただし、

[0058] [数5]

$$\underset{n}{Min}[\ ]$$

は、pを変化させたときに、括弧内の値が最小となる値を示し、

[0059] [数6]

$$\underset{p}{\arg\min}[\ ]$$

は、pを変化させたときに、括弧内の値が最小となるときのpの値を示す。

[0060] 混合パラメータ判定部106は、探索の空間を削減する為に、素片iにおける選択ベクトル候補$h_{i,j}$をコスト値（C（$h_{i,j}$））に基づいて削減する（ステップS205）。例えば、ビームサーチを用いて、最小コスト値から所定の閾値以上大きいコスト値を持つ選択ベクトル候補を削減するようにすればよい。または、コストの小さい候補から所定の個数の候補のみを残すようにすればよい。

[0061] なお、ステップS205の枝狩り処理は、計算量を削減する為の処理であり、計算量に問題がない場合は、この処理を省いても構わない。

[0062] 以上のステップS201からステップS205までの処理を素片i（i＝1，…，n）について繰り返す。混合パラメータ判定部106は、最終素片i＝nの時の最小コストの選択候補

[0063] [数7]

$$s_n = \underset{j}{\arg\min}\, C(h_{n,j})$$

を選択し、接続元の情報を用いて順次バックトラックを

[0064] [数8]

$$s_{n-1} = B(h_{n,s_n})$$

のように行い、式5を用いて選択ベクトル系列Cを求めることが可能になる。

[0065] [数9]

$$C = C_1, C_2, \cdots, C_n = h_{1,s_1}, h_{2,s_2}, \cdots, h_{n,s_n} \quad (式5)$$

[0066] このようにして得られた選択ベクトル系列Cを用いることにより、実音声パラメータが目標パラメータに類似している場合には、実音声パラメータを使用し、そうでない場合は、目標パラメータを用いることが可能となる。

[0067] パラメータ統合部107は、ステップS102で得られた目標パラメータ系列$T=t_1, t_2, \cdots, t_n$とステップS103で得られた実音声パラメータ系列$U=u_1, u_2, \cdots, u_n$と、ステップS104で得られた選択ベクトル系列$C=C_1, C_2, \cdots, C_n$を用いて、合成パラメータ系列$P=p_1, p_2, \cdots, p_n$を式6を用いて生成する(ステップS105)。

[0068] [数10]

$$p_i = C_i \bullet u_i + (1 - C_i) \bullet t_i \quad (式6)$$

[0069] 波形生成部108は、ステップS105により生成された合成パラメータ系列$P=p_1, p_2, \cdots, p_n$を用いて合成音を合成する(ステップS106)。合成方法は特に限定するものではない。目標パラメータ生成部が生成するパラメータにより決定される合成方法を用いればよく、例えば、特許文献2の励振源生成と合成フィルタとを用いて合成音を合成するように構成すればよい。

[0070] 以上のように構成した音声合成装置によれば、目標パラメータを生成する目標パラメータ生成部と、目標パラメータに基づいて実音声パラメータを選択する素片選択部と、目標パラメータと実音声パラメータとの類似度に基づいて、目標パラメータおよび実音声パラメータを切替える選択ベクトル系列Cを生成する混合パラメータ判定部とを用いることにより、実音声パラメータが目標パラメータに類似している場合には、実音声パラメータを使用し、そうでない場合は、目標パラメータを用いることが可能となる。

[0071] 以上のような構成によれば、目標パラメータ生成部102が生成するパラメータの形式と、音声素片DB103が保持する素片の形式とが同一である。そのため、図7に示すように、従来の波形接続型音声合成では目標パラメータとの類似度が低い場合(すなわち、目標パラメータに近い音声素片が音声素片DB103に保持されていない場合)でも、目標パラメータに部分的に近い音声素片を選択し、その音声素片のパラメータのうち、目標パラメータと類似していないパラメータについては、目標パラメータ自体を使用することにより、実音声パラメータを使用していたことによる局所的な音声品質の劣化を防止することが可能となる。

[0072] また、同時に、従来の統計モデルによる音声合成方式では、目標パラメータに類似した素片が存在する場合においても、統計モデルにより生成されるパラメータを用いていた為、肉声感が低下していたが、実音声パラメータを使用することにより(すなわち、目標パラメータに近い音声素片を選択し、その音声素片のパラメータのうち、目標パラメータと類似するパラメータについては、音声素片のパラメータ自体を使用することにより)、肉声感が低下することなく、肉声感が高く高音質な合成音を得ることが可能となる。したがって、目標パラメータによる安定した音質と、実音声パラメータによる肉声感の高い高音質とを両立させた合成音を生成することが可能となる。

[0073] なお、本実施の形態において、選択ベクトルC_iはパラメータのそれぞれの次元毎に設定するように構成したが、図11に示すように全ての次元において同じ値とすることにより、素片iについて、目標パラメータを使用するか、実音声パラメータを使用するかを選択するように構成しても良い。図11には、実音声パラメータを使用する素片の領域601および603と、目標パラメータを使用する素片の領域602および604とが一例として示されている。

[0074] 1つの声質(例えば読上げ調)だけではなく、「怒り」「喜び」等といった多数の声質の合成音を生成する場合には、本発明は非常に効果的である。

[0075] なぜならば、多種多様な声質の音声データをそれぞれ十分な分量用意することは、非常にコストが掛かることから、困難である。

[0076] 上記の説明ではHMMモデルと音声素片とは特に限定していなかったが、HMMモデルと音声素片とを次のように構成することにより、多数の声質の合成音を生成す

ることが可能となる。すなわち、図12に示すように、目標パラメータ生成部102の他に目標パラメータを生成する為に文章HMM作成部302を用意し、文章HMM作成部302が参照するHMMモデル301を標準音声DBとして、通常の読み上げ音声DB101により作成しておく。更に、文章HMM作成部302が、「怒り」「喜び」等の感情音声DB1102により、当該感情を前記HMMモデル301に適応させる。なお、文章HMM作成部302は、特殊な感情を有する音声の統計モデルを作成する統計モデル作成手段に対応する。

[0077] これにより、目標パラメータ生成部102は、感情を有する目標パラメータを生成することができる。適応させる方法は特に限定するものではなく、例えば、橘誠、外4名、"HMM音声合成におけるモデル補間・適応による発話スタイルの多様性の検討"、信学技報 TECHNICAL REPORT OF IEICE SP2003－80(2003－08)に記載の方法により適応することが可能である。また、一方で、素片選択部104が選択する音声素片DBとして前記感情音声DB1102を用いる。

[0078] このように構成することによって、感情音声DB1102により適応されたHMM301を用いて安定した音質で、指定された感情の合成パラメータを生成でき、且つ、素片選択部104により感情音声DB1102から、感情音声素片を選択する。混合パラメータ判定部106により、HMMにより生成されたパラメータと、感情音声DB1102から選択されたパラメータとの混合を判定し、パラメータ統合部107により統合する。

[0079] 従来の波形重畳型の感情を表現する音声合成装置は、十分な音声素片DBを用意しなければ、高音質な合成音を生成することが困難であった。また、従来のHMM音声合成では、モデル適応は可能であるが、統計処理であるので合成音になまり(肉声感の低下)が生じるという問題があった。しかし、上記のように感情音声DB1102をHMMモデルの適用データおよび音声素片DBとして構成することにより、適応モデルにより生成される目標パラメータによる安定した音質と、感情音声DB1102から選択される実音声パラメータによる高品質で肉声感の高い音質とを両立した合成音声を生成することが可能なる。つまり、目標パラメータに類似した実音声パラメータが選択できた場合には、従来は、統計モデルにより生成される肉声感が低いパラメータを使用していたのに対して、実音声パラメータを使用することにより、肉声感が高く、

且つ自然な感情を含む音質を実現できる。一方、目標パラメータとの類似度が低い実音声パラメータが選択された場合には、従来の波形接続型音声合成方式では、局所的に音質が劣化していたのに対し、目標パラメータを使用することにより、局所的な劣化を防ぐことが可能となる。

[0080] したがって、本発明によれば、複数の声質の合成音を作成したい場合においても、それぞれの声質で大量の音声を収録することなく、かつ、統計モデルにより生成される合成音よりも肉声感の高い合成音を生成することが可能となる。

[0081] また、感情音声DB1102の変わりに、特定の人物による音声DBを用いることにより、特定の個人に適応した合成音を同様に生成することが可能である。

[0082] (実施の形態2)

図13は、本発明の実施の形態2の音声合成装置の構成図である。図13において、図4と同じ構成要素については同じ符号を用い、説明を省略する。

[0083] 図13において、目標パラメータパターン生成部801は、目標パラメータ生成部102で生成された目標パラメータに基づいて、後述する目標パラメータパターンを生成する処理部である。

[0084] 音声素片DB103A1～103C2は、音声素片DB103の部分集合であり、目標パラメータパターン生成部801により生成された目標パラメータパターンそれぞれに対応したパラメータを格納する音声素片DBである。

[0085] 素片選択部104A1～104C2は、目標パラメータパターン生成部801により生成された目標パラメータパターンに最も類似した素片を音声素片DB103A1～103C2からそれぞれ選択する処理部である。

[0086] 以上のように音声合成装置を構成することにより、パラメータパターンごとに選択した音声素片のパラメータの部分集合を組み合わせることができる。これにより、単一の素片に基づいて選択した場合と比較して、目標パラメータにより類似した実音声に基づくパラメータを生成することが可能となる。

[0087] 以下に、本発明の実施の形態2の音声合成装置の動作について図14のフローチャートを用いて説明する。

[0088] 言語解析部101は、入力されたテキストを言語的に解析し、発音記号およびアクセ

ント記号を生成する(ステップS101)。目標パラメータ生成部102は、発音記号およびアクセント記号に基づいて、上述のHMM音声合成法により、再合成可能なパラメータ系列$T=t_1, t_2, \cdots, t_n$を生成する(ステップS102)。このパラメータ系列を目標パラメータと呼ぶ。

[0089] 目標パラメータパターン生成部801は、目標パラメータを図15に示すようなパラメータの部分集合に分割する(ステップS301)。分割の方法は特に限定するものではないが、例えば以下のように分割することが可能である。なお、これらの分け方は一例であり、これらに限定されるものではない。

[0090] ・音源情報と声道情報

・基本周波数とスペクトル情報と揺らぎ情報

・基本周波数と音源スペクトル情報と声道スペクトル情報と音源揺らぎ情報

[0091] このようにして分割したパラメータパターンを複数用意する(図15のパターンA、パターンB、パターンC)。図15では、パターンAを、パターンA1, A2およびA3の3つの部分集合に分割している。また、同様にパターンBを、パターンB1およびB2の2つの部分集合に分割しており、パターンCを、パターンC1およびC2の2つの部分集合に分割している。

[0092] 次に、素片選択部104A1～104C2は、ステップS301で生成された複数のパラメータパターンのそれぞれについて、素片選択を行なう(ステップS103)。

[0093] ステップS103では、素片選択部104A1～104C2は、目標パラメータパターン生成部801によって生成されたパターンの部分集合(パターンA1、A2、…、C2)毎に最適な音声素片を音声素片DB103A1～103C2から選択し、素片候補集合列Uを作成する。各素片候補u_iの選択の方法は、上記実施の形態1と同じ方法でよい。

[0094] [数11]

$$U = U_1, U_2, \cdots, U_n$$
$$U_i = (u_{i1}, u_{i2}, \cdots, u_{im}) \quad \text{(式7)}$$

[0095] 図13では、素片選択部および音声素片DBは複数用意されているが、物理的に用意する必要はなく、実施の形態1の音声素片DBおよび素片選択部を複数回使用するように設計しても良い。

[0096] 組み合わせ判定部802は、それぞれの素片選択部(A1, A2, ・・・, C2)により選択された実音声パラメータの組み合わせベクトル系列Sを決定する(ステップS302)。組み合わせベクトル系列Sは式8のように定義する。

[0097] [数12]

$$S = S_1, S_2, \cdots, S_n$$
$$S_i = (s_1, s_2, \cdots, s_m) \qquad \text{(式8)}$$
$$s_i = \begin{cases} 0 : \text{i番目の部分集合を採用しない場合} \\ 1 : \text{i番目の部分集合を採用する場合} \end{cases}$$

[0098] 組み合わせベクトルの決定方法(ステップS302)について図16を用いて詳しく説明する。探索アルゴリズムを図16のフローチャートを用いて説明する。素片i(i=1, …, n)に対して、ステップS401からステップS405の処理が順次繰り返される。

[0099] 組み合わせ判定部802は、対象となる素片に対して、組み合わせベクトルS_iの候補h_iとして、p個の候補$h_{i,1}$, $h_{i,2}$, …, $h_{i,p}$を生成する(ステップS401)。生成する方法は特に限定するものではない。例えば図17A(a)および図17B(a)に示すように、ある一つのパターンに含まれる部分集合のみを生成しても良い。また、図17A(b)および図17B(b)に示すように、複数のパターンに属する部分集合をパラメータ同士(907と908)で、重なりが生じないように生成しても良い。また、図17A(c)および図17B(c)のパラメータ909に示すように、複数のパターンに属する部分集合をパラメータ同士で一部重なりが生じるように生成しても良い。この場合は、重なりが生じたパラメータに関しては、それぞれのパラメータの重心点を用いるようにする。また、図17A(d)および図17B(d)のパラメータ910に示すように、複数のパターンに属する部分集合をパラメータ同士を組み合わせた時に、一部パラメータが欠落した状態になるように生成しても良い。この場合は、欠落したパラメータに関しては、目標パラメータ生成部によって生成された目標パラメータで代用する。

[0100] ターゲットコスト判定部105aは、選択ベクトルS_iの候補$h_{i,1}$, $h_{i,2}$, …, $h_{i,p}$と、素片iの目標パラメータt_iとの類似度に基づくコストを式9により計算する(ステップS402)。

[0101] [数13]

$$TargetCost(h_{i,j}) = \omega_1 \times Tc(h_{i,j} \bullet U_i, t_i) \qquad \text{(式9)}$$

[0102] ここで、ω_1は、重みである。重みの決定方法は特に限定するものではないが、経験に基づき決定することが可能である。また、$h_{i,j}$、$\cdot U_i$は、ベクトル$h_{i,j}$とベクトルU_iの内積であり、組み合わせベクトル$h_{i,j}$、によって決定される各素片候補の部分集合を示す。関数Tcは、パラメータ間の類似度に基づくコスト値を算出する。算出方法は特に限定するものではないが、例えば、各パラメータ次元間の差分の重み付け加算により算出することが可能である。

[0103] 連続性判定部105bは、選択ベクトル候補$h_{i,j}$、それぞれについて、1つ前の選択ベクトル候補との連続性に基づくコストを式10を用いて評価する(ステップS403)。

[0104] [数14]

$$ContCost(h_{i,j}, h_{i-1,r}) = Cc(h_{i,j} \bullet U_i, h_{i-1,r} \bullet U_{i-1}) \quad \text{(式10)}$$

[0105] 関数Ccは、2つの素片パラメータの連続性に基づくコストを評価する関数である。算出方法は特に限定するものではないが、例えば、素片i－1の最終フレームと素片iの先頭フレームにおける各パラメータ次元の差分値の重み付け和により計算すればよい。

[0106] 組み合わせ判定部802は、選択ベクトル候補$h_{i,j}$、に対するコスト(C($h_{i,j}$))を算定し、同時に素片i－1に対する選択ベクトル候補$h_{i-1,r}$、のうちどの選択ベクトル候補と接続すべきかを示す接続元(B($h_{i,j}$))を式11に基づいて決定する(ステップS404)。

[0107] [数15]

$$C(h_{i,j}) = \mathrm{TargetCost}(h_{i,j}) + \min_p \left[\mathrm{ContCost}(h_{i,j}, h_{i-1,p}) + C(h_{i-1,p})\right]$$
$$B(h_{i,j}) = \operatorname{argmin}_p \left[\mathrm{ContCost}(h_{i,j}, h_{i-1,p}) + C(h_{i-1,p})\right] \quad \text{(式11)}$$

[0108] 組み合わせ判定部802は、探索の空間を削減する為に、素片iにおける選択ベクトル候補$h_{i,j}$、をコスト値(C($h_{i,j}$))に基づいて削減する(ステップS405)。例えば、ビームサーチを用いて、最小コスト値から所定の閾値以上大きいコスト値を持つ選択ベクトル候補を削減するようにすればよい。または、コストの小さい候補から所定の個数の候補のみを残すようにすればよい。

[0109] なお、ステップS405の枝狩り処理は、計算量を削減する為のステップであり、計算

量に問題がない場合は、処理を省いても構わない。

[0110] 以上のステップS401からステップS405までの処理を素片i(i＝1, …, n)について繰り返す。組み合わせ判定部802は、最終素片i＝nの時の最小コストの選択候補

[0111] [数16]

$$s_n = \underset{j}{\operatorname{argmin}} C(h_{n,j})$$

を選択する。以降は、接続元の情報を用いて順次バックトラックを

[0112] [数17]

$$s_{n-1} = B(h_{n,s_n})$$

のように行い、式12により組み合わせベクトル系列Sを求めることが可能になる。

[0113] [数18]

$$S = S_1, S_2, \cdots, S_n = h_{1,s_1}, h_{2,s_2}, \cdots, h_{n,s_n} \quad (式12)$$

[0114] パラメータ統合部107は、組み合わせ判定部802により決定された組み合わせベクトルに基づいて、各素片選択部(A1, A2, ・・・, C2)により選択された素片のパラメータを式13を用いて統合する(ステップS105)。図18は、統合の例を示す図である。この例では、素片1の組み合わせベクトルS_1＝(A_1, 0, 0, 0, 0, 0, C_2)であり、パターンAによるA1と、パターンCによるC2の組み合わせが選択されている。これにより、パターンA1により選択された素片1501と、パターンC2により選択された素片1502を組み合わせて素片1のパラメータとしている。以下、S_2, …, S_nまで繰り返すことにより、パラメータ系列を得ることが可能である。

[0115] [数19]

$$p_i = S_i \bullet U_i \quad (式13)$$

[0116] 波形生成部108は、パラメータ統合部107により生成された合成パラメータに基づいて合成音を合成する(ステップS106)。合成方法は特に限定するものではない。

[0117] 以上のように構成した音声合成装置によれば、目標パラメータ生成部が生成する

目標パラメータに近いパラメータ系列を、複数の実音声素片の部分集合である実音声パラメータを組み合わせる。これによって、図18に示すように、従来の波形接続型音声合成方式では目標パラメータとの類似度が低い実音声パラメータが選択された場合には、局所的に音質が劣化していたの対し、目標パラメータとの類似度が低い場合には、複数のパラメータ集合ごとに選択された複数の実音声素片の実音声パラメータを組み合わせることにより、目標パラメータに類似した実音声パラメータを合成することが可能となる。これにより安定して目標パラメータに近い素片を選択することが可能となり、かつ実音声素片を用いている為、高音質となる。つまり、高音質と安定性の双方を両立させた合成音を生成することが可能となる。

[0118] 特に、素片DBが十分に大きくない場合においても、音質と安定性を両立した合成音を得ることが可能となる。なお、本実施の形態において、1つの声質(例えば読上げ調)だけではなく、「怒り」「喜び」等といった多数の声質の合成音を生成する場合には、図12に示すように、目標パラメータ生成部102が目標パラメータを生成する為に文章HMM作成部302を用意し、文章HMM作成部302が参照するHMMモデルを標準音声DBとして、通常の読み上げ音声DB1101により作成しておく。更に、「怒り」「喜び」等の感情音声DB1102により、前記HMMモデル301を適応する。適応する方法は特に限定するものではなく、例えば、「橘誠外4名、"HMM音声合成におけるモデル補間・適応による発話スタイルの多様性の検討"、信学技報 TECHNICAL REPORT OF IEICE SP2003－80(2003－08)」に記載の方法により適応することが可能である。また、一方で、素片選択部104が選択する音声素片DBとして前記感情音声DB1102を用いる。

[0119] このように構成することによって、感情音声DB1102により適応されたHMM301を用いて安定した音質で、指定された感情の合成パラメータを生成でき、且つ、素片選択部104により感情音声DB1102から、感情音声素片を選択する。混合パラメータ判定部により、HMMにより生成されたパラメータと、感情音声DB1102から選択されたパラメータとの混合を判定し、パラメータ統合部107により統合する。これにより、従来の感情を表現する音声合成装置は、十分な音声素片DBを用意しなければ、高音質な合成音を生成することが困難であったのに対し、感情音声DB1102を音声素片

DBとして用いた場合においても、複数のパラメータ集合ごとに選択された複数の実音声素片の実音声パラメータを組み合わせる。これにより目標パラメータに類似した実音声パラメータに基づくパラメータにより高品質な音質とを両立した合成音声を生成することが可能なる。

[0120] また、感情音声DB1102の変わりに、別人による音声DBを用いることにより、個人に適応した合成音を同様に生成することが可能である。

[0121] また、言語解析部101は必ずしも必須の構成要件ではなく、言語解析された結果である発音記号やアクセント情報等が音声合成装置に入力されるような構成であっても構わない。

[0122] なお、本実施の形態1および2に示した音声合成装置をLSI(集積回路)で実現することも可能である。

[0123] 例えば、実施の形態1に係る音声合成装置をLSI(集積回路)で実現すると、言語解析部101、目標パラメータ生成部102、素片選択部104、コスト算出部105、混合パラメータ判定部106、パラメータ統合部107、波形生成部108のすべてを1つのLSIで実現することができる。または、各処理部を1つのLSIで実現することもできる。さらに、各処理部を複数のLSIで構成することもできる。音声素片DB103は、LSIの外部の記憶装置により実現してもよいし、LSIの内部に備えられたメモリにより実現してもよい。LDIの外部の記憶装置により音声素片DB103を実現する場合には、インターネット経由で音声素片DB103に記憶されている音声素片を取得しても良い。

[0124] ここでは、LSIとしたが、集積度の違いにより、IC、システムLSI、スーパーLSI、ウルトラLSIと呼称されることもある。

[0125] また、集積回路化の手法はLSIに限られるものではなく、専用回路または汎用プロセサにより実現してもよい。LSI製造後に、プログラムすることが可能なFPGA(Field Programmable Gate Array)や、LSI内部の回路セルの接続や設定を再構成可能なリコンフィギュラブル・プロセッサを利用しても良い。

[0126] さらには、半導体技術の進歩又は派生する別技術によりLSIに置き換わる集積回路化の技術が登場すれば、当然、その技術を用いて音声合成装置を構成する処理部の集積化を行ってもよい。バイオ技術の適応等が可能性としてありえる。

[0127] また、本実施の形態1および2に示した音声合成装置をコンピュータで実現することも可能である。図19は、コンピュータの構成の一例を示す図である。コンピュータ1200は、入力部1202と、メモリ1204と、CPU1206と、記憶部1208と、出力部1210とを備えている。入力部1202は、外部からの入力データを受け付ける処理部であり、キーボード、マウス、音声入力装置、通信I／F部等から構成される。メモリ1204は、プログラムやデータを一時的に保持する記憶装置である。CPU1206は、プログラムを実行する処理部である。記憶部1208は、プログラムやデータを記憶する装置であり、ハードディスク等からなる。出力部1210は、外部にデータを出力する処理部であり、モニタやスピーカ等からなる。

[0128] 例えば、実施の形態1に係る音声合成装置をコンピュータ1200で実現した場合には、言語解析部101、目標パラメータ生成部102、素片選択部104、コスト算出部105、混合パラメータ判定部106、パラメータ統合部107、波形生成部108は、CPU1206上で実行されるプログラムに対応し、音声素片DB103は、記憶部1208に記憶される。また、CPU1206で計算された結果は、メモリ1204や記憶部1208に一旦記憶される。メモリ1204や記憶部1208は、言語解析部101等の各処理部とのデータの受け渡しに利用されてもよい。また、音声合成装置をコンピュータに実行させるためのプログラムは、フロッピー（登録商標）ディスク、CD－ROM、DVD－ROM、不揮発性メモリ等に記憶されていてもよいし、インターネットを経由してコンピュータ1200のCPU1206に読み込まれてもよい。

[0129] 今回開示された実施の形態はすべての点で例示であって制限的なものではないと考えられるべきである。本発明の範囲は上記した説明ではなくて特許請求の範囲によって示され、特許請求の範囲と均等の意味および範囲内でのすべての変更が含まれることが意図される。

産業上の利用可能性

[0130] 本発明にかかる音声合成装置は、実音声による高音質の特徴と、モデルベース合成の安定性を有し、カーナビゲーションシステムや、ディジタル家電のインタフェース等として有用である。また、音声DBを用いてモデル適応を行うことにより声質を変更が可能な音声合成装置等の用途にも応用できる。

請求の範囲

[1] 少なくとも発音記号を含む情報から、音声を合成することが可能なパラメータ群である目標パラメータを素片単位で生成する目標パラメータ生成部と、

予め録音された音声を、前記目標パラメータと同じ形式のパラメータ群からなる音声素片として素片単位で記憶している音声素片データベースと、

前記目標パラメータに対応する音声素片を前記音声素片データベースより選択する素片選択部と、

音声素片ごとに、前記目標パラメータのパラメータ群および前記音声素片のパラメータ群を統合してパラメータ群を合成するパラメータ群合成部と、

合成された前記パラメータ群に基づいて、合成音波形を生成する波形生成部とを備える

ことを特徴とする音声合成装置。

[2] 前記パラメータ群合成部は、

前記素片選択部により選択された音声素片の部分集合と当該音声素片の部分集合に対応する前記目標パラメータの部分集合とに基づいて、当該音声素片の部分集合を選択することによるコストまたは当該目標パラメータの部分集合を選択することによるコストを算出するコスト算出部と、

前記コスト算出部によるコスト値に基づいて、前記目標パラメータと前記音声素片との最適なパラメータの組み合わせを、素片単位ごとに判定する混合パラメータ判定部と、

前記混合パラメータ判定部により判定された組み合わせに基づいて、前記目標パラメータと前記音声素片とを統合することによりパラメータ群を合成するパラメータ統合部とを有する

ことを特徴とする請求項1に記載の音声合成装置。

[3] 前記コスト算出部は、

前記素片選択部により選択された音声素片の部分集合と当該音声素片の部分集合に対応する前記目標パラメータの部分集合との非類似性を示すコストを算出するターゲットコスト判定部を有する

ことを特徴とする請求項2に記載の音声合成装置。

[4] 前記コスト算出部は、さらに、

前記素片選択部により選択された音声素片の部分集合を当該音声素片の部分集合に対応する前記目標パラメータの部分集合に置き換えた音声素片に基づいて、時間的に連続する音声素片同士の不連続性を示すコストを算出する連続性判定部を有する

ことを特徴とする請求項3に記載の音声合成装置。

[5] 前記音声素片データベースは、

標準的な感情を有する音声素片を記憶している標準音声データベースと、

特殊な感情を有する音声素片を記憶している感情音声データベースとを有し、

前記音声合成装置は、さらに、前記標準的な感情を有する音声素片および前記特殊な感情を有する音声素片に基づいて、特殊な感情を有する音声の統計モデルを作成する統計モデル作成手段を備え、

前記目標パラメータ生成部は、前記特殊な感情を有する音声の統計モデルに基づいて、目標パラメータを素片単位で生成し、

前記素片選択部は、前記目標パラメータに対応する音声素片を前記感情音声データベースより選択する

ことを特徴とする請求項1に記載の音声合成装置。

[6] 前記パラメータ群合成部は、

前記目標パラメータ生成部により生成された目標パラメータを、少なくとも1つ以上の部分集合に分割することによって得られるパラメータパターンを少なくとも1つ以上生成する目標パラメータパターン生成部と、

前記目標パラメータパターン生成部により生成された前記目標パラメータの部分集合ごとに、当該部分集合に対応する音声素片を前記音声素片データベースより選択する素片選択部と、

前記素片選択部により選択された音声素片の部分集合と当該音声素片の部分集合に対応する前記目標パラメータの部分集合とに基づいて、当該音声素片の部分集合を選択することによるコストを算出するコスト算出部と、

前記コスト算出部によるコスト値に基づいて、前記目標パラメータの部分集合の最適な組み合わせを、素片ごとに判定する組み合わせ判定部と、

前記組み合わせ判定部により判定された組み合わせに基づいて、前記素片選択部により選択された前記音声素片の部分集合を統合することによりパラメータ群を合成するパラメータ統合部とを有する

ことを特徴とする請求項1に記載の音声合成装置。

[7] 前記組み合わせ判定部は、前記音声素片の部分集合を組み合わせる際に、部分集合同士に重なりが生じる場合には、重なりが生じたパラメータに関しては平均値を当該パラメータの値として、最適な組み合わせを判定する

ことを特徴とする請求項6に記載の音声合成装置。

[8] 前記組み合わせ判定部は、前記音声素片の部分集合を組み合わせる際に、パラメータの欠落が生じる場合には、欠落したパラメータを目標パラメータにより代用して、最適な組み合わせを判定する

ことを特徴とする請求項6に記載の音声合成装置。

[9] 少なくとも発音記号を含む情報から、音声を合成することが可能なパラメータ群である目標パラメータを素片単位で生成するステップと、

前記目標パラメータに対応する音声素片を、予め録音された音声を前記目標パラメータと同じ形式のパラメータ群からなる音声素片として素片単位で記憶している音声素片データベースより選択するステップと、

音声素片ごとに、前記目標パラメータのパラメータ群および前記音声素片のパラメータ群を統合してパラメータ群を合成するステップと、

合成された前記パラメータ群に基づいて、合成音波形を生成するステップとを含む

ことを特徴とする音声合成方法。

[10] 少なくとも発音記号を含む情報から、音声を合成することが可能なパラメータ群である目標パラメータを素片単位で生成するステップと、

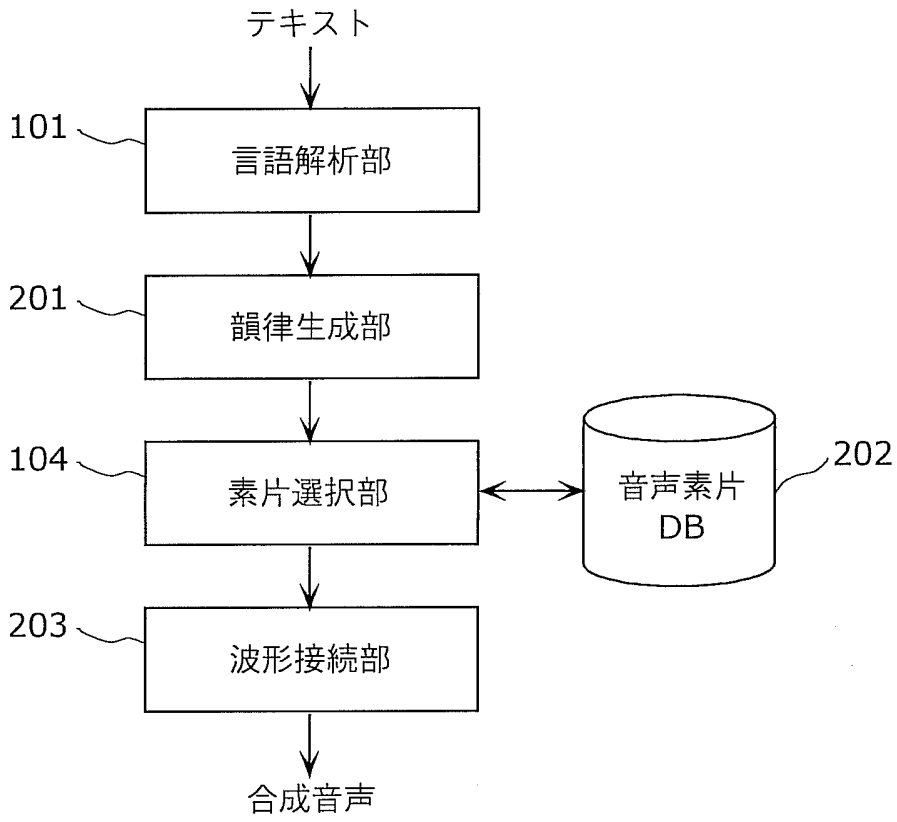
前記目標パラメータに対応する音声素片を、予め録音された音声を前記目標パラメータと同じ形式のパラメータ群からなる音声素片として素片単位で記憶している音声素片データベースより選択するステップと、

音声素片ごとに、前記目標パラメータのパラメータ群および前記音声素片のパラメータ群を統合してパラメータ群を合成するステップと、

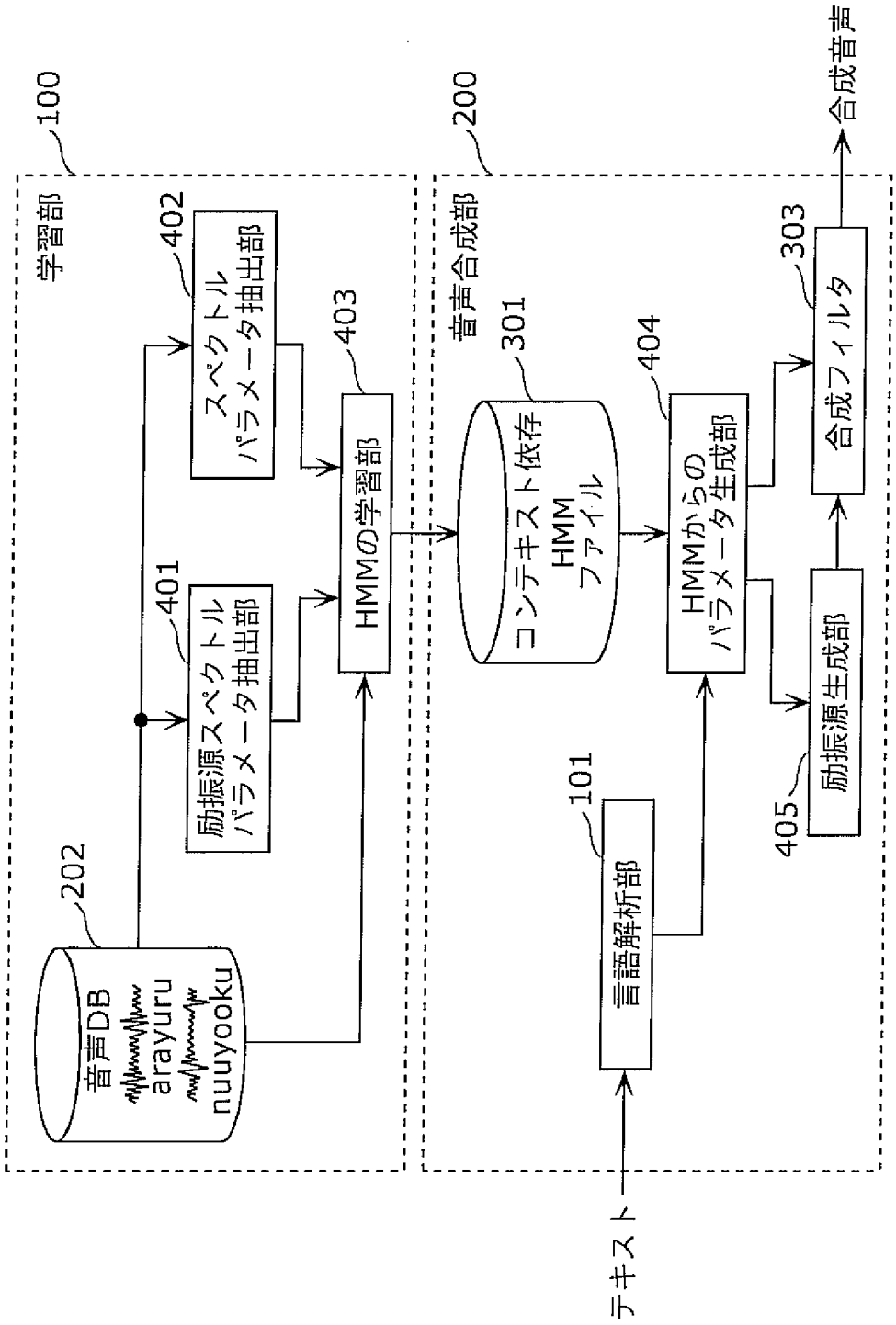
合成された前記パラメータ群に基づいて、合成音波形を生成するステップとをコンピュータに実行させる

ことを特徴とするプログラム。

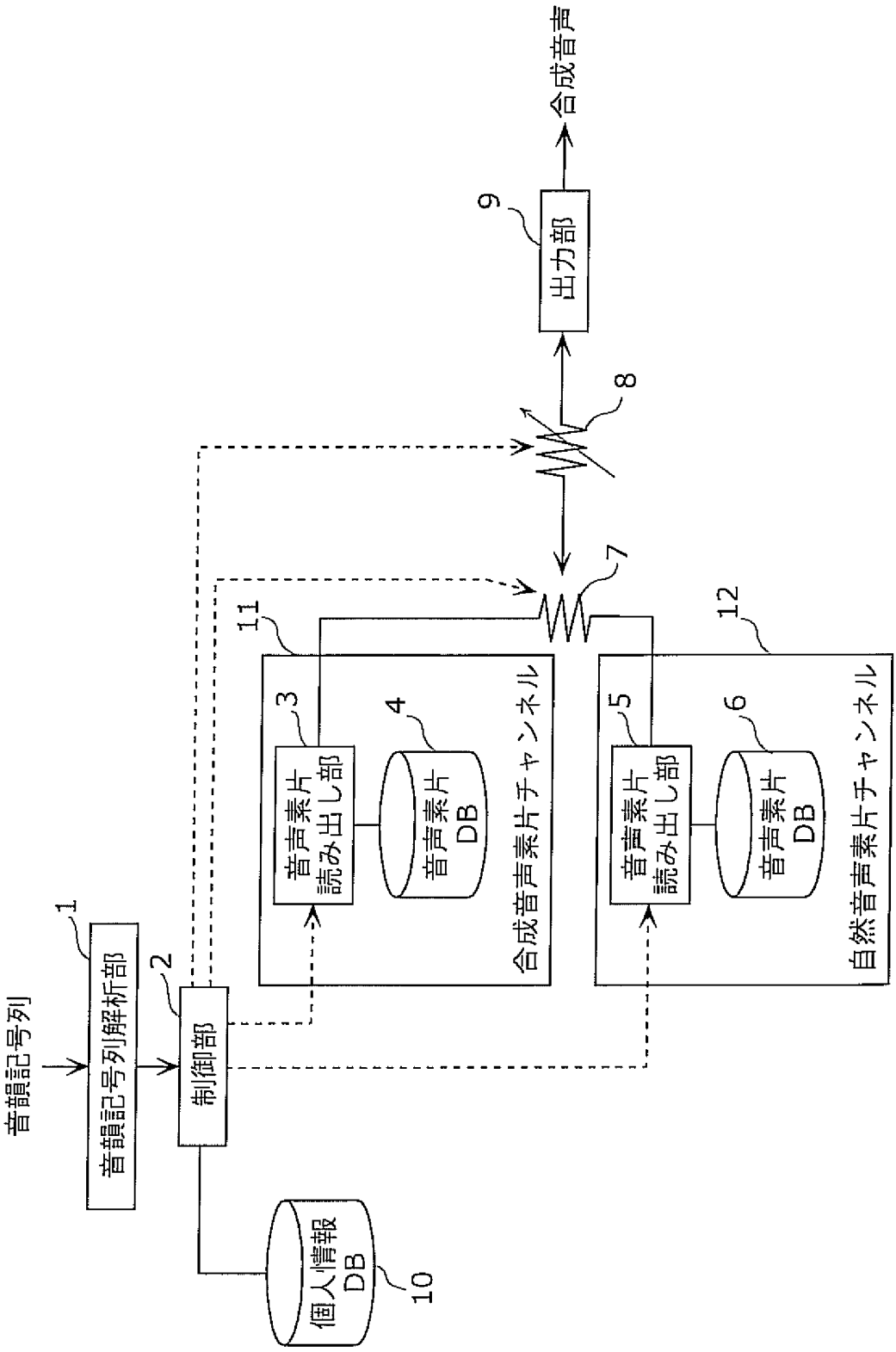
[図1]



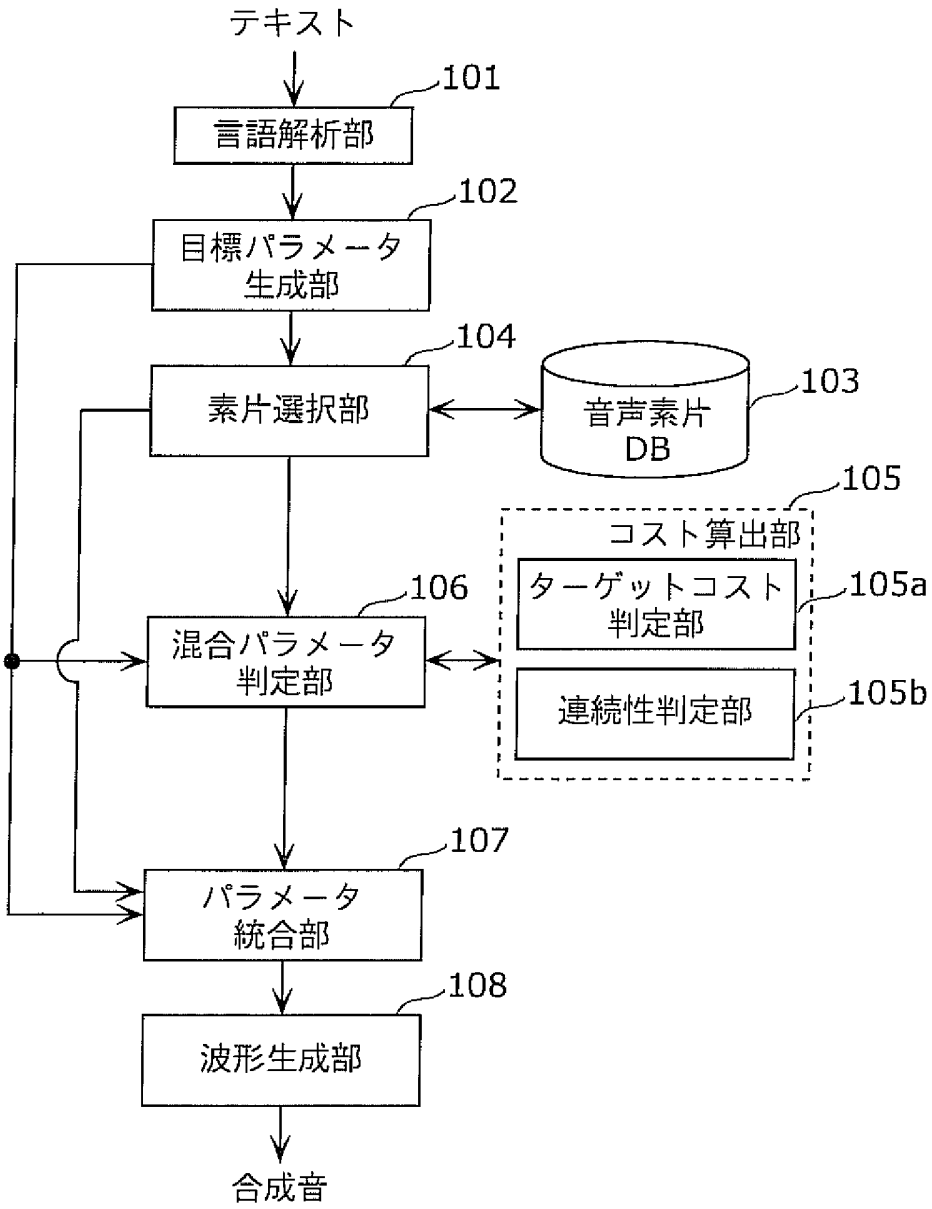
[図2]


100
学習部
202
音声DB
arayuru
nuuyooku
401
励振源スペクトル
パラメータ抽出部
402
スペクトル
パラメータ抽出部
403
HMMの学習部
200
音声合成部
301
コンテキスト依存
HMM
ファイル
404
HMMからの
パラメータ生成部
303
合成フィルタ
合成音声
101
言語解析部
405
励振源生成部
テキスト

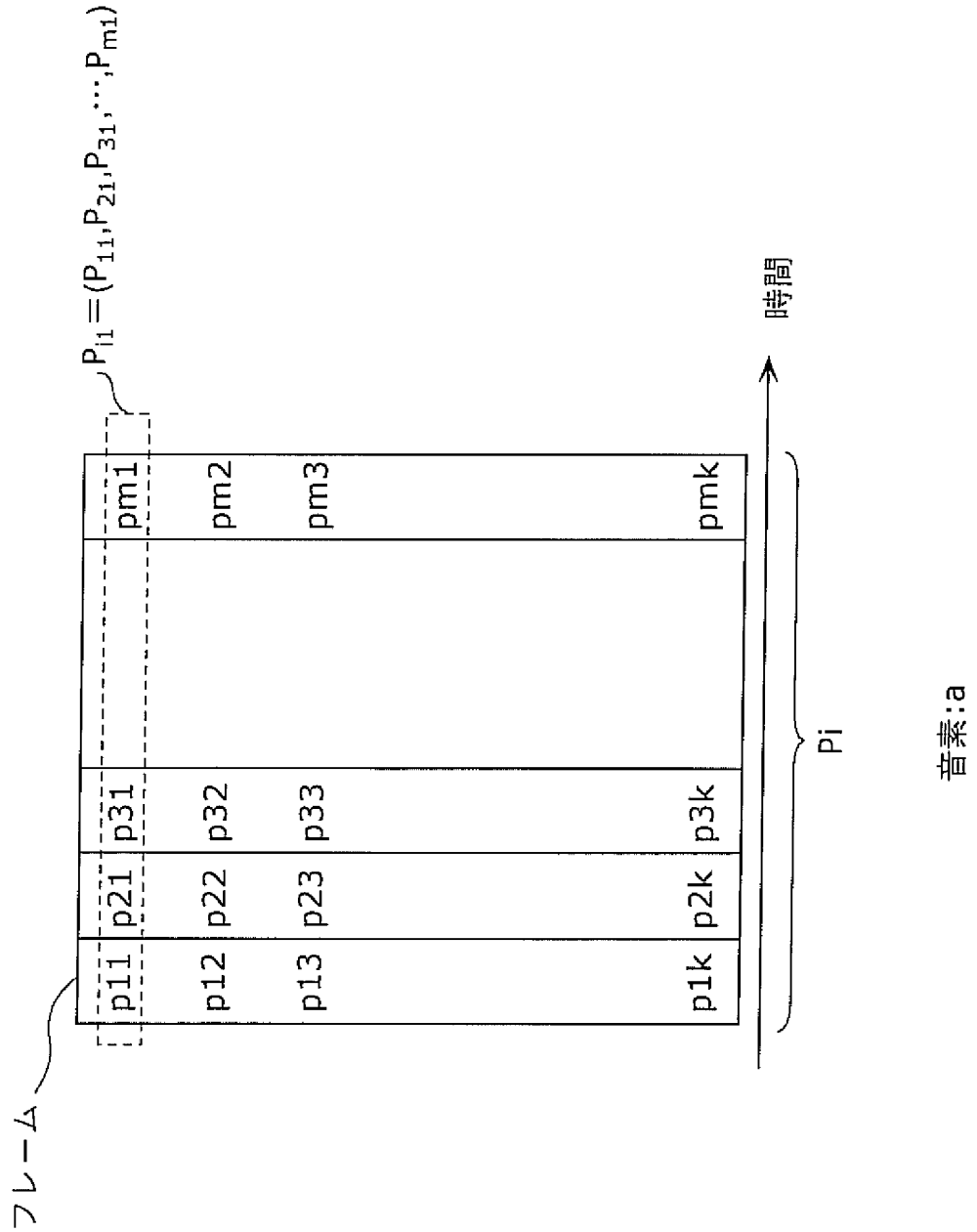
[図3]


音韻記号列
1
音韻記号列解析部
2
制御部
個人情報DB
10
合成音素片チャンネル
11
音声素片読み出し部
3
音声素片DB
4
自然音声素片チャンネル
12
音声素片読み出し部
5
音声素片DB
6
7
8
出力部
9
合成音声

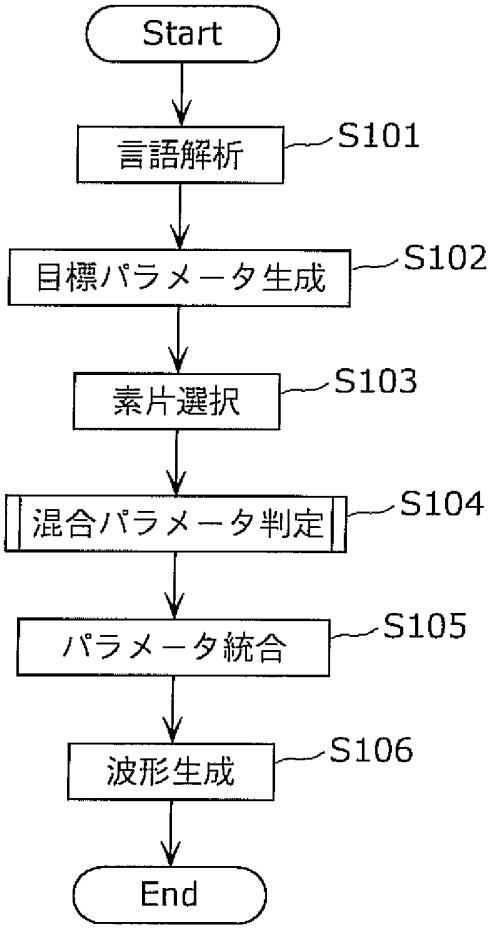
[図4]


テキスト
101
言語解析部
102
目標パラメータ
生成部
104
素片選択部
103
音声素片
DB
105
コスト算出部
ターゲットコスト
判定部
105a
106
混合パラメータ
判定部
連続性判定部
105b
107
パラメータ
統合部
108
波形生成部
合成音

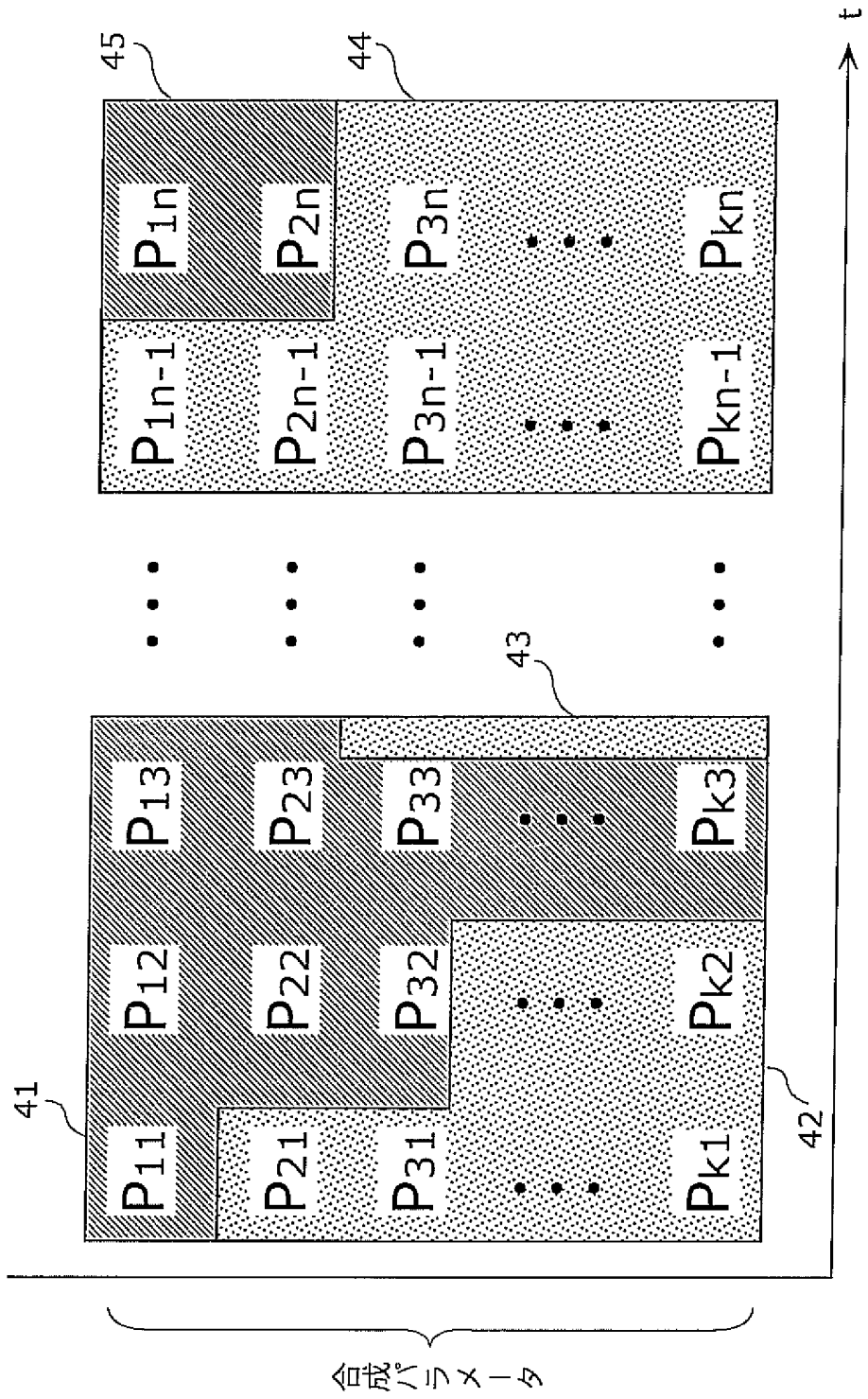
[図5]


$P_{i1}=(P_{11},P_{21},P_{31},\cdots,P_{m1})$
フレーム
p11
p21
p31
pm1
p12
p22
p32
pm2
p13
p23
p33
pm3
p1k
p2k
p3k
pmk
時間
Pi
音素:a

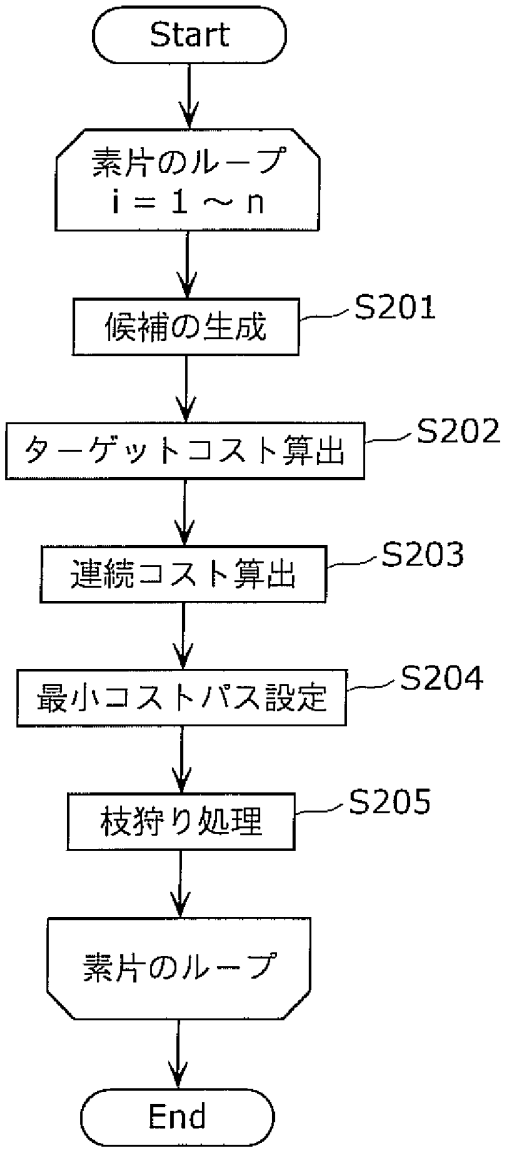
[図6]


Start
言語解析
S101
目標パラメータ生成
S102
素片選択
S103
混合パラメータ判定
S104
パラメータ統合
S105
波形生成
S106
End

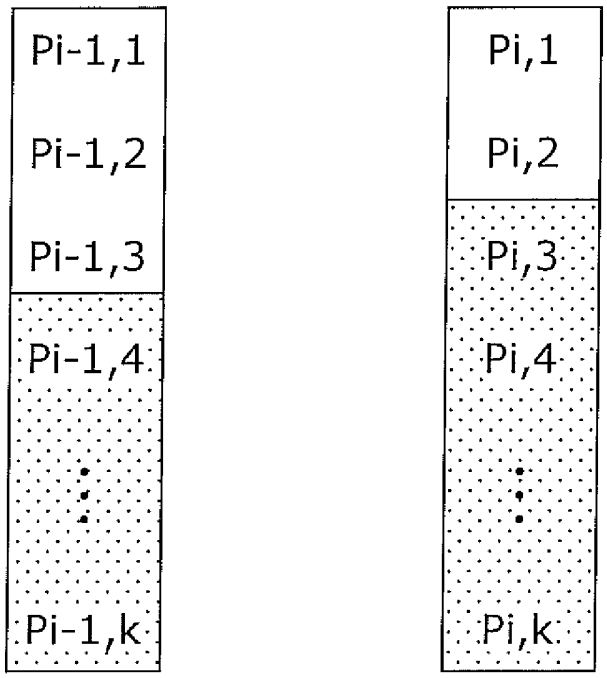
[図7]


45
44
43
41
42
t
P1n
P2n
P3n
Pkn
P1n-1
P2n-1
P3n-1
Pkn-1
P13
P23
P33
Pk3
P12
P22
P32
Pk2
P11
P21
P31
Pk1
合成パラメータ

[図8]


Start
素片のループ
i = 1 ～ n
候補の生成
S201
ターゲットコスト算出
S202
連続コスト算出
S203
最小コストパス設定
S204
枝狩り処理
S205
素片のループ
End

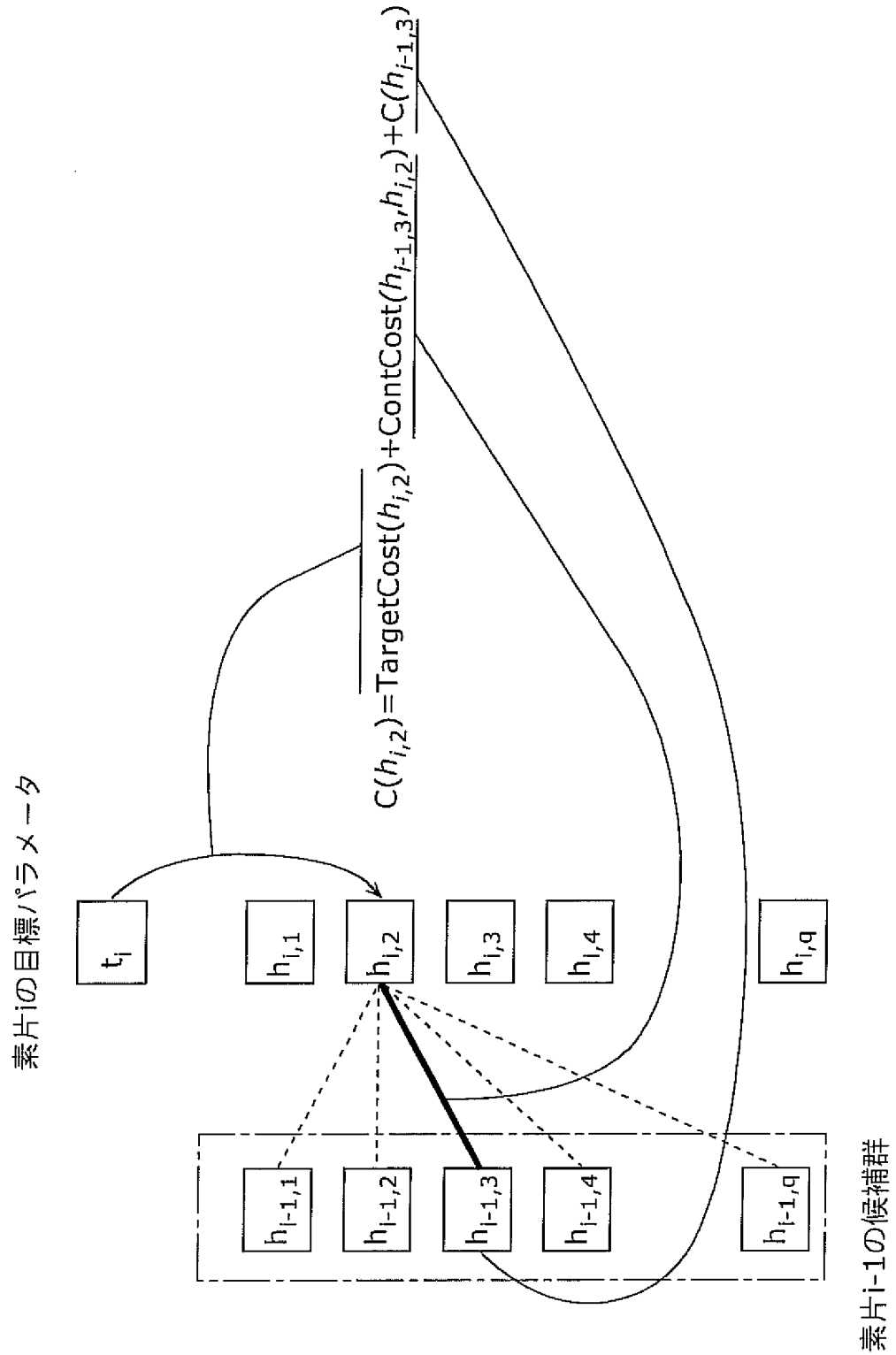
[図9]



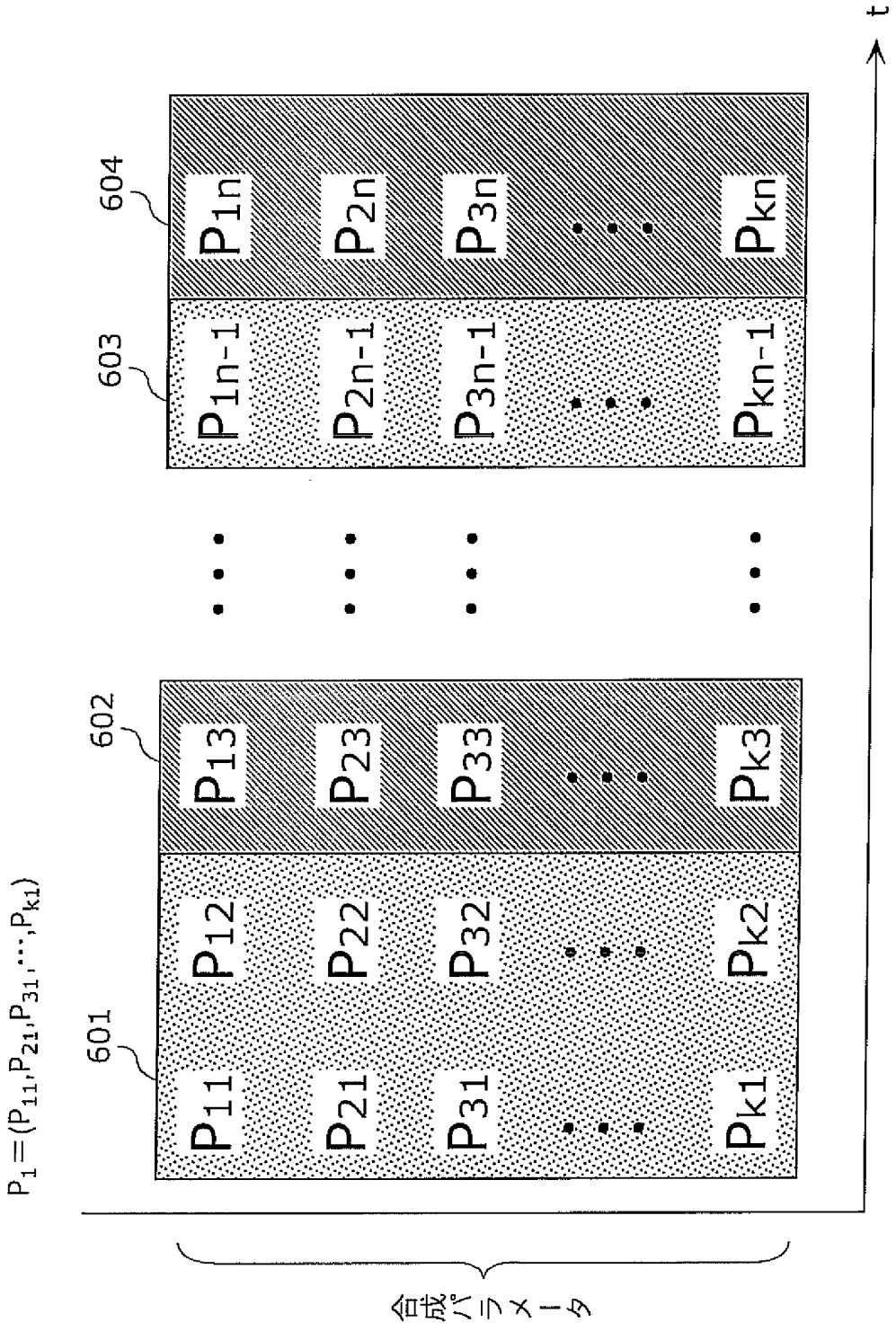
Ci-1=(1,1,1,0,0,0 0) Ci=(1,1,0,0,0,0 0)

(Ci-1とCiの差分=1) ＜ 所定の閾値

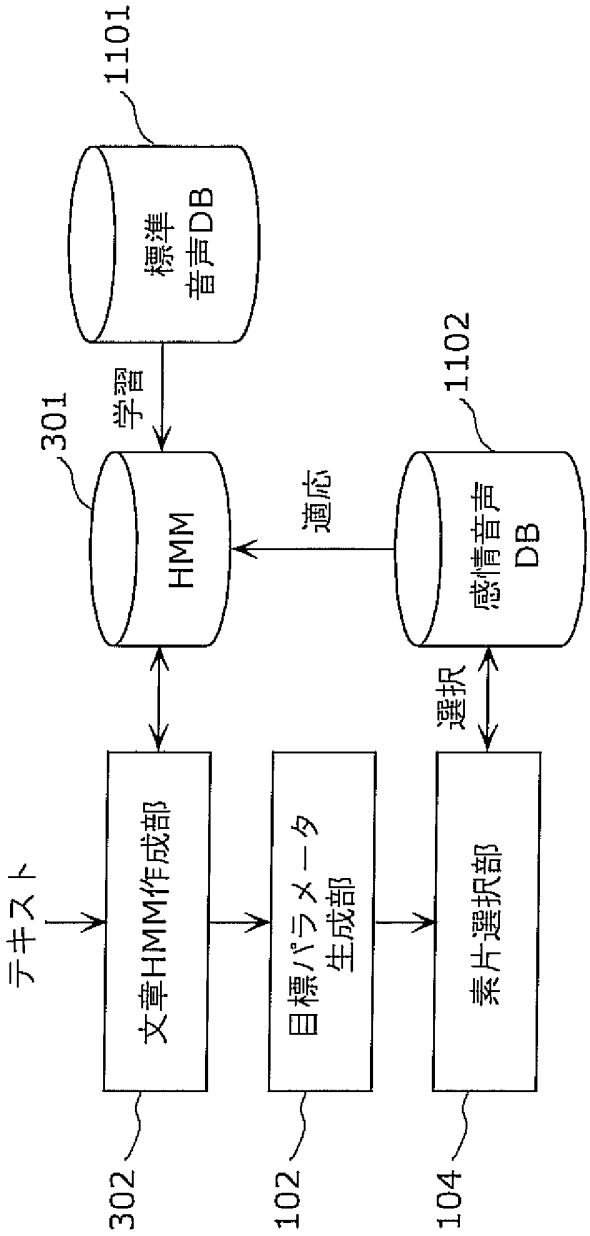
[図10]


素片iの目標パラメータ
t_i
$h_{i,1}$
$h_{i,2}$
$h_{i,3}$
$h_{i,4}$
$h_{i,q}$
$C(h_{i,2})=\mathrm{TargetCost}(h_{i,2})+\mathrm{ContCost}(h_{i-1,3},h_{i,2})+C(h_{i-1,3})$
$h_{i-1,1}$
$h_{i-1,2}$
$h_{i-1,3}$
$h_{i-1,4}$
$h_{i-1,q}$
素片i-1の候補群

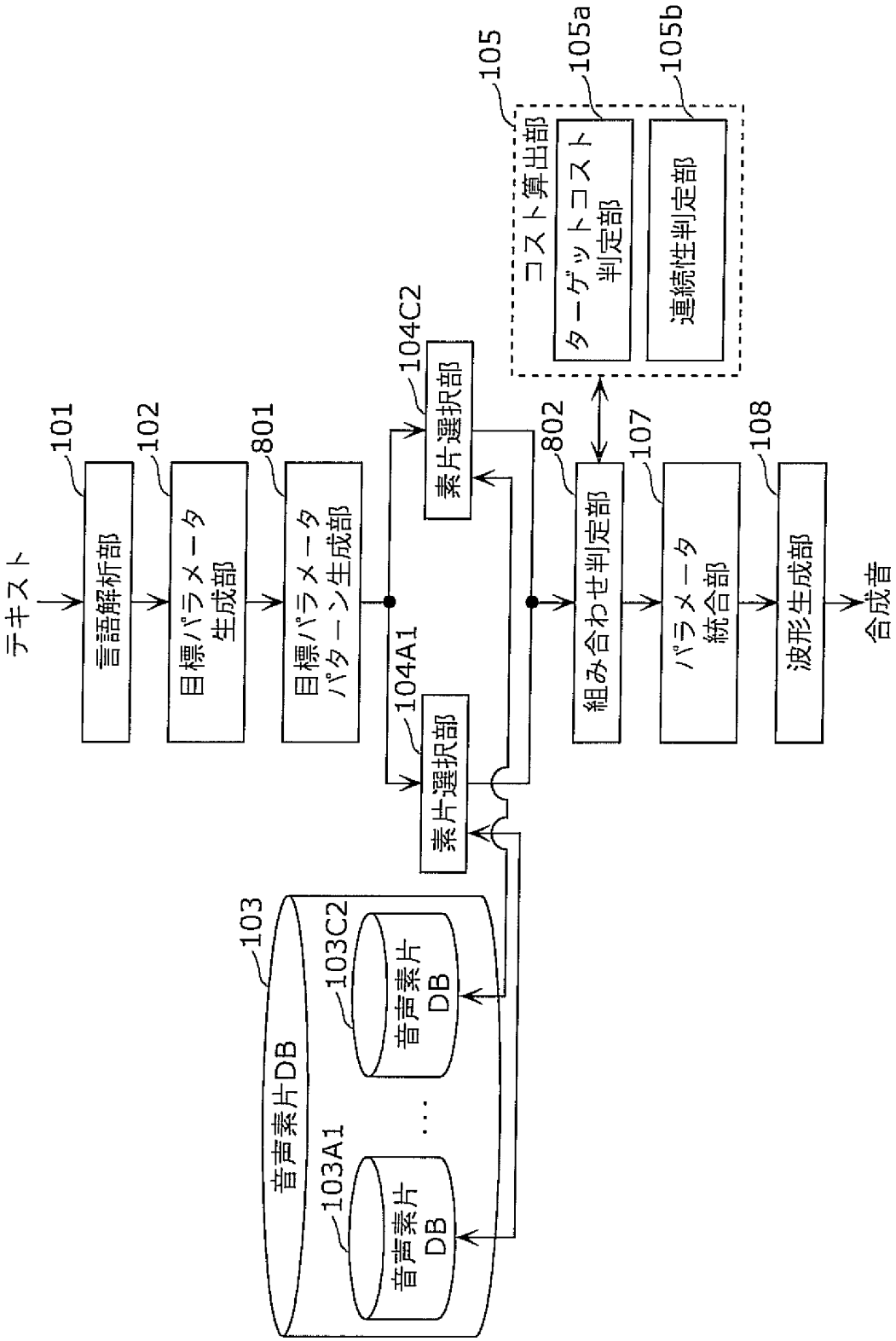
[図11]


t
604
603
602
601
P_{1n} P_{2n} P_{3n} ・・・ P_{kn}
P_{1n-1} P_{2n-1} P_{3n-1} ・・・ P_{kn-1}
P_{13} P_{23} P_{33} ・・・ P_{k3}
P_{12} P_{22} P_{32} ・・・ P_{k2}
P_{11} P_{21} P_{31} ・・・ P_{k1}
$P_1=(P_{11},P_{21},P_{31},\cdots,P_{k1})$
各成分データ

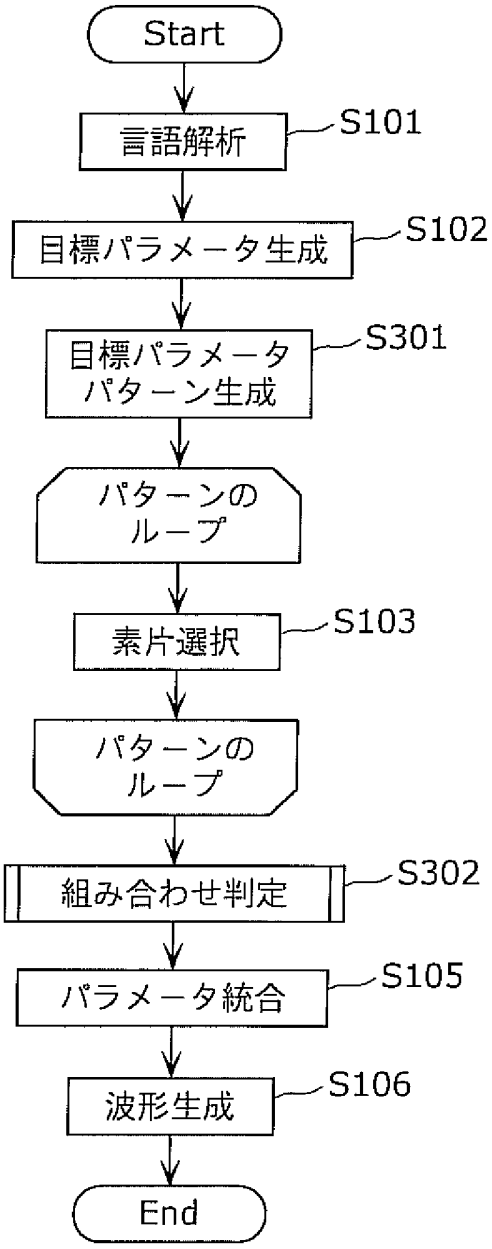
[図12]


1101
標準
音声DB
学習
301
HMM
適応
1102
感情音声
DB
選択
テキスト
文章HMM作成部
目標パラメータ
生成部
素片選択部
302
102
104

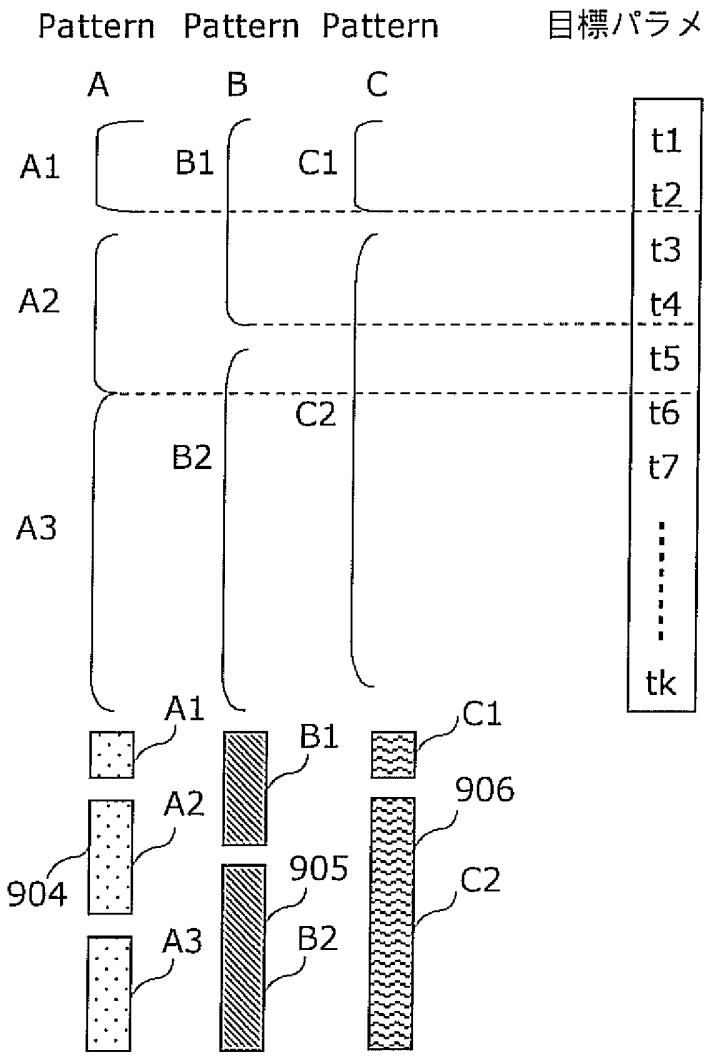
[図13]


テキスト
101
言語解析部
102
目標パラメータ生成部
801
目標パラメータパターン生成部
104C2
素片選択部
104A1
素片選択部
103
音声素片DB
103C2
音声素片DB
103A1
音声素片DB
105
コスト算出部
105a
ターゲットコスト判定部
105b
連続性判定部
802
組み合わせ判定部
107
パラメータ統合部
108
波形生成部
合成音

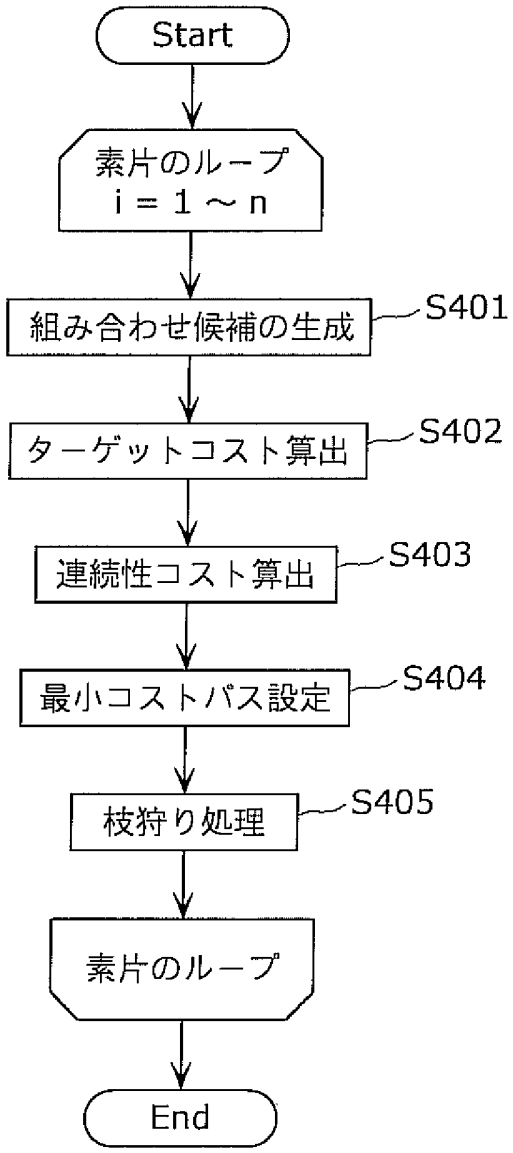
[図14]


Start
言語解析
S101
目標パラメータ生成
S102
目標パラメータ
パターン生成
S301
パターンの
ループ
素片選択
S103
パターンの
ループ
組み合わせ判定
S302
パラメータ統合
S105
波形生成
S106
End

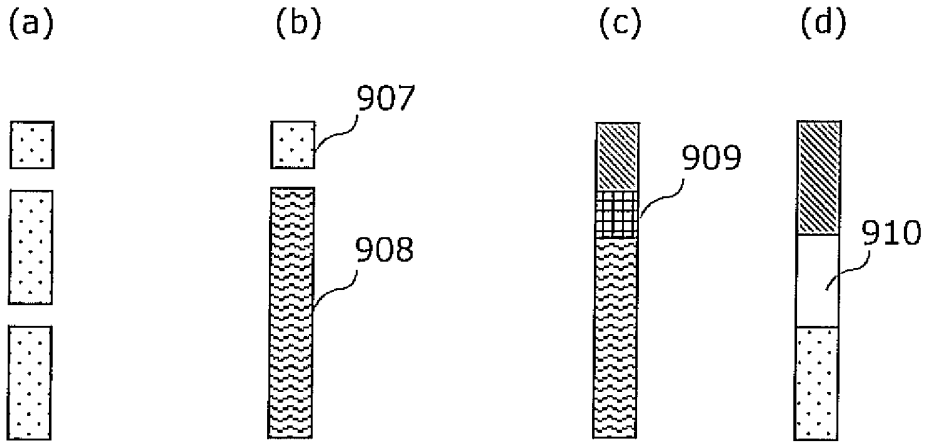
[図15]


Pattern
Pattern
Pattern
目標パラメータ
A
B
C
A1
B1
C1
A2
B2
C2
A3
t1
t2
t3
t4
t5
t6
t7
tk
A1
B1
C1
A2
906
904
905
C2
A3
B2

[図16]


Start
素片のループ
i = 1 ～ n
組み合わせ候補の生成
S401
ターゲットコスト算出
S402
連続性コスト算出
S403
最小コストパス設定
S404
枝狩り処理
S405
素片のループ
End

[図17A]


(a)
(b)
(c)
(d)
907
908
909
910

[図17B]

S = (A1, A2, A3, B1, B2, C1, C2) の場合

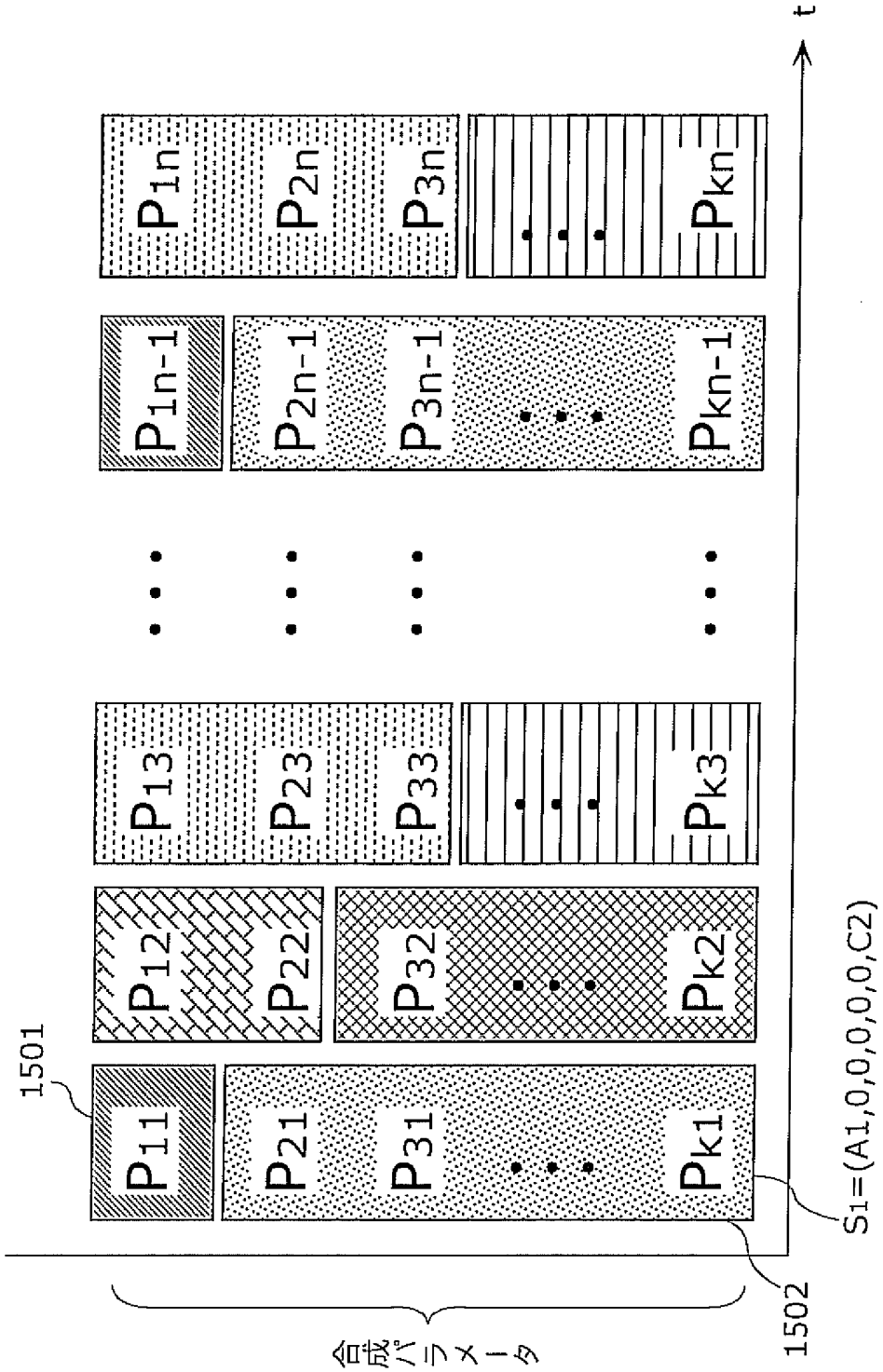
(a) S = (1, 1, 1, 0, 0, 0, 0)

(b) S = (1, 0, 0, 0, 1, 0, 0)

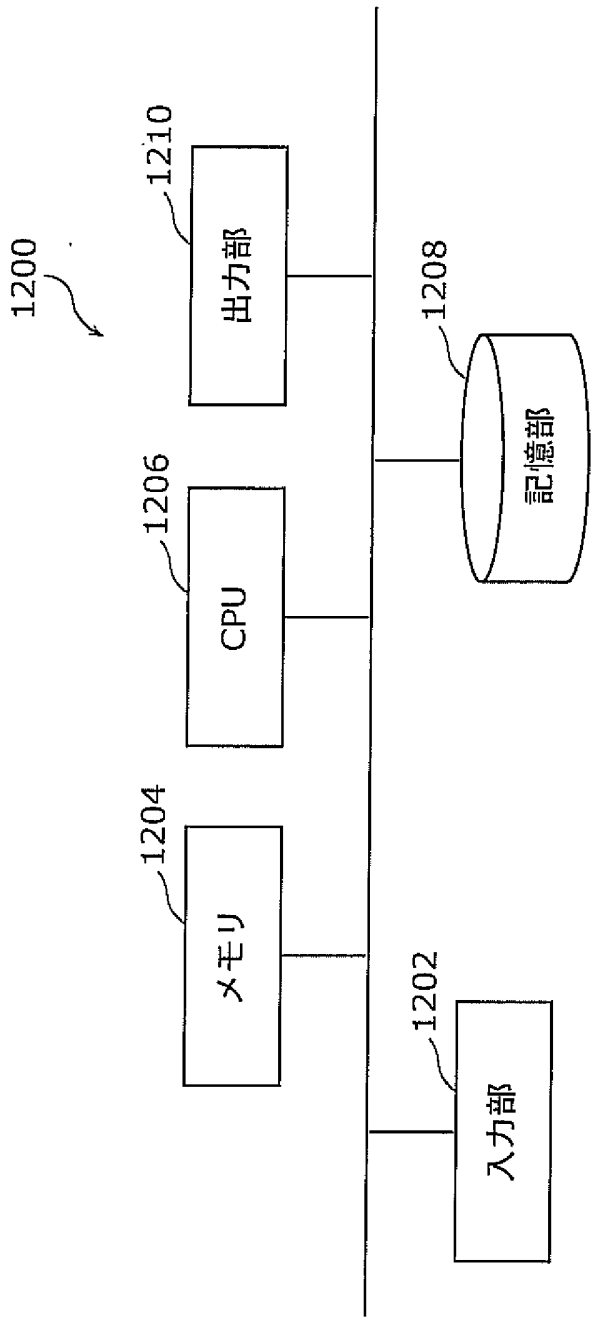
(c) S = (0, 0, 0, 0, 1, 0, 1)

(d) S = (0, 0, 1, 0, 0, 0, 1)

[図18]


t
P1n
P2n
P3n
Pkn
P1n-1
P2n-1
P3n-1
Pkn-1
P13
P23
P33
Pk3
P12
P22
P32
Pk2
P11
P21
P31
Pk1
1501
S1=(A1,0,0,0,0,0,C2)
1502
合成パラメータ

[図19]



INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2006/309288

A. CLASSIFICATION OF SUBJECT MATTER
G10L13/08(2006.01)

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)
G10L13/00-13/08(2006.01)

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 1996-2006
Kokai Jitsuyo Shinan Koho 1971-2006 Toroku Jitsuyo Shinan Koho 1994-2006

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
JSTPlus(JDream2)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2003-295880 A (Fujitsu Ltd.), 15 October, 2003 (15.10.03), Full text; all drawings & US 2003/0187651 A1	1-10
A	Toshimitsu MINOWA et al., "Inritsu no Vector o Riyo shita Onsei Gosei Hoshiki", National Institute of Advanced Industrial Science and Technology, 19 May, 2000 (19.05.00), Vol.100, No.97, SP2000-4, pages 25 to 31	1-10
A	Toshiyuki SANO et al., "Onso Setsuzokugata Onsei Gosei to Yokuyo Henkan Gijutsu no Yugo ni yoru Shizen na Gosei Onsei no Kakutoku", OMRON TECHNICS, 15 January, 2000 (15.01.00), Vol.39, No.4, pages 324 to 329	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	Date of mailing of the international search report
06 June, 2006 (06.06.06)	13 June, 2006 (13.06.06)
Name and mailing address of the ISA/ Japanese Patent Office Facsimile No.	Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2006/309288

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 2000-181476 A (Toyota Motor Corp.), 30 June, 2000 (30.06.00), Full text; all drawings (Family: none)	1-10
A	JP 8-63187 A (Fujitsu Ltd.), 08 March, 1996 (08.03.96), Full text; all drawings (Family: none)	1-10
A	JP 5-61498 A (Ricoh Co., Ltd.), 12 March, 1993 (12.03.93), Full text; all drawings (Family: none)	1-10

<Concerning the object of search>

Claims 1, 5, 9, 10 fail to specify when and how the parameter group of target parameters and the parameter group of speech fragments are "integrated". Therefore, the inventions of claims 1, 5, 9, 10 involve a wide scope of technical matter.

Claims 2, 6 fail to specify how the "costs" calculated by a cost calculating section are determined. Therefore, the inventions of claims 2, 6 involve a wide scope of "costs". Further, claims 2, 6 fail to specify how the "costs" are used and how the parameter group of target parameters and the parameter group of speech fragments are "integrated". Therefore, the inventions of claims 2, 6 involve a wide scope of technical matter.

However, the matter supported by the disclosure of the description within the meaning of PCT Article 6 is a parameter group "integration" in which the "cost" required to judge for each dimension of a synthesization parameter whether or not "the target parameter" is similar to "the real voice parameter" most approximate to the target parameter is determined, the "real voice parameter" is used for the parameter of the dimension for which the cost is judged to be low (similar), and the "target parameter" is used for the parameter of the dimension for which the cost is judged to be high (not similar).

Consequently, the search has been conducted on the scope supported by the disclosure of description, that is, a speech synthesizer having "a parameter group generating section" carrying out the "integration".

国際調査報告

国際出願番号　ＰＣＴ／ＪＰ２００６／３０９２８８

Ａ．　発明の属する分野の分類（国際特許分類（ＩＰＣ））

Int.Cl.　*G10L13/08(2006.01)*

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（ＩＰＣ））

Int.Cl.　G10L13/00-13/08(2006.01)

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	１９２２－１９９６年
日本国公開実用新案公報	１９７１－２００６年
日本国実用新案登録公報	１９９６－２００６年
日本国登録実用新案公報	１９９４－２００６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(JDream2)

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	関連する請求の範囲の番号
A	JP 2003-295880 A (富士通株式会社) 2003.10.15, 全文，全図 & US 2003/0187651 A1	1-10
A	蓑輪利光 他，韻律のベクトルを利用した音声合成方式， 電子情報通信学会技術研究報告，2000.05.19, Vol.100, No.97, SP2000-4, p.25-31	1-10

☑　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	国際調査報告の発送日
０６．０６．２００６	１３．０６．２００６

国際調査機関の名称及びあて先
日本国特許庁（ＩＳＡ／ＪＰ）
郵便番号１００－８９１５
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）　５Ｚ　３３５２
荏原　雄一
電話番号　０３－３５８１－１１０１　内線　３５４１

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	関連する請求の範囲の番号
A	佐野敏幸 他，音素接続型音声合成と抑揚変換技術の融合による自然な合成音声の獲得，OMRON TECHNICS，2000.01.15，Vol.39，No.4，p.324-329	1-10
A	JP 2000-181476 A（トヨタ自動車株式会社）2000.06.30，全文，全図　（ファミリーなし）	1-10
A	JP 8-63187 A（富士通株式会社）1996.03.08，全文，全図　（ファミリーなし）	1-10
A	JP 5-61498 A（株式会社リコー）1993.03.12，全文，全図　（ファミリーなし）	1-10

＜調査の対象について＞
　請求の範囲１、５、９、１０では、目標パラメータのパラメータ群と、音声素片のパラメータ群とを、どのような場合に、そして、どのように「統合」するのかが特定されていないため、請求の範囲１、５、９、１０に係る発明は、広範囲なものを包含する。
　また、請求の範囲２及び６では、コスト算出部によって算出される「コスト」が、それぞれどのように求められるコストであるのか特定されていないため、広範囲な「コスト」を包含する。さらに、上記「コスト」をどのように利用し、目標パラメータのパラメータ群と、音声素片のパラメータ群とを、どのように「統合」するのかが特定されていないため、請求の範囲２及び６に係る発明は、広範囲なものを包含する。

　しかしながら、ＰＣＴ第６条の意味において明細書の開示により裏付けられているのは、合成パラメータの次元ごとに、「目標パラメータ」と、該目標パラメータに最も近い「実音声のパラメータ」とが、類似しているかどうかの「コスト」を求め、コストが小さい（類似している）と判定された次元のパラメータには前記「実音声のパラメータ」を、コストが大きい（類似していない）と判定された次元のパラメータには前記「目標パラメータ」を用いる、というパラメータ群の「統合」処理である。
　よって、調査は、明細書の開示により裏付けられている範囲、すなわち、上記「統合」処理を行なう「パラメータ群合成部」を備えた音声合成装置について行なった。