(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 9 月 30 日(30.09.2010)

PCT

(10) 国際公開番号
WO 2010/110095 A1

(51) 国際特許分類:
G10L 13/06 (2006.01) *G10L 21/04* (2006.01)

(21) 国際出願番号: PCT/JP2010/054250

(22) 国際出願日: 2010 年 3 月 12 日(12.03.2010)

(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 2009-074707 2009 年 3 月 25 日(25.03.2009) 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 (米国についてのみ): 森中 亮 (MORINAKA, Ryo) [JP/JP]. 籠嶋 岳彦 (KAGOSHIMA, Takehiko) [JP/JP].

(74) 代理人: 鈴江 武彦, 外 (SUZUYE, Takehiko et al.); 〒1050001 東京都港区虎ノ門１丁目１２番９号 鈴榮特許綜合事務所内 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, TH, 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, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

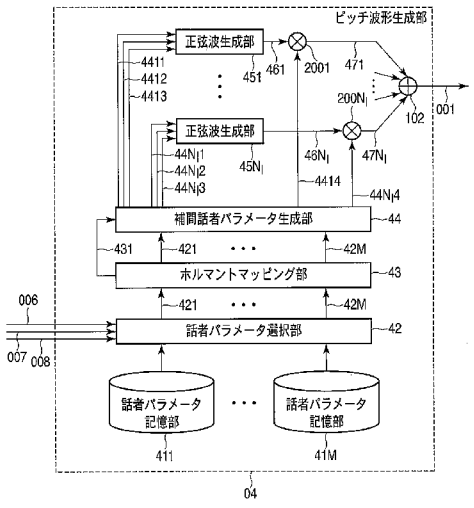
添付公開書類:
— 国際調査報告（条約第 21 条(3)）

WO 2010/110095 A1

(54) **Title**: VOICE SYNTHESIZER AND VOICE SYNTHESIZING METHOD

(54) 発明の名称 : 音声合成装置及び音声合成方法

[図3]



04 PITCH WAVEFORM GENERATION UNIT
42 SPEAKER PARAMETER SELECTION UNIT
43 FORMANT MAPPING UNIT
44 INTERPOLATED SPEAKER PARAMETER GENERATION UNIT
411 SPEAKER PARAMETER STORAGE UNIT
41M SPEAKER PARAMETER STORAGE UNIT
45N1 SINE WAVE GENERATION UNIT
451 SINE WAVE GENERATION UNIT

(57) **Abstract**: A voice synthesizing apparatus comprises a selection unit (42) which selects a speaker parameter prepared for each pitch waveform corresponding to a voice of a speaker and including a formant frequency, a formant phase, a formant power, and a window function regarding each of a plurality of formants contained in each pitch waveform, one by one for each speaker, to obtain a plurality of speaker parameters (421, …, 42M); a mapping unit (43) for correlating the formants among the plurality of speaker parameters using a cost function on the basis of the formant frequency and formant power; and a generation unit (44) for generating an interpolated speaker parameter by interpolating the formant frequency, the formant phase, the formant power, and the window function, in accordance with a desired interpolation ratio, among the formants correlated by the mapping unit (43).

(57) 要約: 音声合成装置は、話者の音声に相当するピッチ波形毎に用意され、各ピッチ波形に含まれる複数のホルマントの各々に関するホルマント周波数、ホルマント位相、ホルマントパワー及び窓関数を含む話者パラメータを話者毎に１つずつ選択して、複数の話者パラメータ（４２１，・・・，４２Ｍ）を得る選択部（４２）と、ホルマント周波数及び前記ホルマントパワーに基づくコスト関数を利用して複数の話者パラメータの間でホルマント同士の対応付けを行うマッピング部（４３）と、マッピング部（４３）によって互いに対応付けられているホルマント同士でホルマント周波数、ホルマント位相、ホルマントパワー及び窓関数を所望の補間比率に従って補間して補間話者パラメータを生成する生成部（４４）とを具備する。

明　細　書

発明の名称 ： 音声合成装置及び音声合成方法

技術分野

[0001] 本発明は、テキスト音声合成に関する。

背景技術

[0002] 任意の文章（テキスト）を表す音声信号を人工的に生成する技術をテキスト音声合成という。テキスト音声合成は、言語処理、韻律処理及び音声信号合成処理の３段階処理によって実現される。

[0003] 第１段階となる言語処理において、入力されたテキストに対する形態素解析、構文解析などが行われる。次に、第２段階となる韻律処理において、上記言語処理結果に基づきアクセント、イントネーションに関する処理が行われ、音韻系列（音韻記号列）及び韻律情報（基本周波数、音韻継続時間長、パワーなど）が出力される。そして、第３段階となる音声信号合成処理において音韻系列及び韻律情報に基づき音声信号が合成される。

[0004] ある種のテキスト音声合成の基本原理は、音声素片(speech segment)と呼ばれる特徴パラメータを接続することである。具体的には、音声素片は、ＣＶ、ＣＶＣ、ＶＣＶなど（尚、Ｃは子音、Ｖは母音を表す）の比較的短い音声の特徴パラメータを指す。予め用意されている音声素片を、ピッチ及び継続時間長を制御して接続することにより、任意の音韻記号列を合成することができる。このようなテキスト音声合成において、利用可能な音声素片の品質が合成音声の品質に強い影響を与える。

[0005] 特許文献１記載の音声合成方法は、音声素片を例えばホルマント(formant)周波数を用いて表現している。そして、この音声合成方法は、１つのホルマントを表す波形（以下、単にホルマント波形と称する）を、ホルマント周波数と等しい周波数の正弦波に対して窓関数を乗じることにより生成し、このホルマント波形を複数重ね合わせ（加算）することにより音声信号を合成する。従って、特許文献１記載の音声合成方法によれば、音韻または声質を直

接的に制御できるので、合成音声の声質を変化させるなどの柔軟な制御を比較的容易に実現できる。

[0006] 特許文献２記載の音声合成装置は、複数の話者の音声スペクトルデータを所定の補間比率を用いて内挿することにより、補間音声スペクトルデータ生成する。従って、特許文献２記載の音声合成装置によれば、比較的簡易な構成であるにも関わらず、合成音声の声質を制御できる。

先行技術文献

特許文献

[0007] 特許文献1：特許第３７３２７９３号公報

特許文献2：特許第２９５１５１４号公報

発明の概要

発明が解決しようとする課題

[0008] 特許文献１記載の音声合成方法は、声の太さを変えるための制御関数を用いて音声素片に含まれる全てのホルマント周波数を変換することにより、ホルマントを高周波側にシフトさせて合成音声の声質を細くしたり、低周波側にシフトさせて合成音声の声質を太くしたりすることができる。しかしながら、特許文献１記載の音声合成方法は、複数の話者に基づく補間音声を合成していない。

[0009] 特許文献２記載の音声合成装置は、複数の話者に基づく補間音声を合成するものの、簡易な構成であるため補間音声の品質は必ずしも高くない。特に、特許文献２記載の音声合成装置は、ホルマント位置（ホルマント周波数）やホルマント数の異なる複数の音声スペクトルデータを補間した場合に十分な品質の補間音声が得られないおそれがある。

[0010] 従って、本発明は所望の声質の補間音声を合成可能な音声合成装置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

[0011] 本発明の一態様に係る音声合成装置は、話者の音声に相当するピッチ波形

毎に用意され、各ピッチ波形に含まれる複数のホルマントの各々に関するホルマント周波数、ホルマント位相、ホルマントパワー及び窓関数を含む話者パラメータを話者毎に１つずつ選択して、複数の話者パラメータを得る選択部と、前記ホルマント周波数及び前記ホルマントパワーに基づくコスト関数を利用して前記複数の話者パラメータの間でホルマント同士の対応付けを行うマッピング部と、前記マッピング部によって互いに対応付けられているホルマント同士でホルマント周波数、ホルマント位相、ホルマントパワー及び窓関数を所望の補間比率に従って補間して補間話者パラメータを生成する生成部と、前記補間話者パラメータを用いて、前記補間比率に基づく補間話者の音声に相当するピッチ波形を合成する合成部とを具備する。

発明の効果

[0012] 本発明によれば、所望の声質の補間音声を合成可能な音声合成装置を提供できる。

図面の簡単な説明

[0013] [図1]第１の実施形態に係る音声合成装置を示すブロック図。

[図2]図１の有声音生成部が行う生成処理を示す図。

[図3]図１のピッチ波形生成部の内部を示すブロック図。

[図4]図３の話者パラメータ記憶部に記憶される話者パラメータの一例を示す図。

[図5]図３の話者パラメータ選択部によって選択される話者パラメータを概念的に示す図。

[図6]図３のホルマントマッピング部が行うマッピング処理を示すフローチャート。

[図7]図６のマッピング処理の開始時におけるマッピング結果の一例を示す図。

[図8]図６のマッピング処理の終了時におけるマッピング結果の一例を示す図。

[図9]図８のマッピング結果に基づく話者Ｘ及び話者Ｙ間のホルマントの対応

関係を示す図。

[図10]図３の補間パラメータ生成部が行う生成処理を示すフローチャート。

[図11]図３のピッチ波形生成部が正弦波及び窓関数に基づいて補間音声に相当するピッチ波形を生成する様子を示す図。

[図12]図３のピッチ波形生成部が正弦波及び窓関数に基づいて補間音声に相当するピッチ波形を生成する様子を示す図。

[図13]第２の実施形態に係る音声合成装置の補間話者パラメータ生成部が行う生成処理を示すフローチャート。

[図14]図１３のステップＳ４５０において行われる挿入処理の詳細を示すフローチャート。

[図15]図１４の処理に基づくホルマントの挿入例を示す図。

[図16]第３の実施形態に係る音声合成装置のピッチ波形生成部を示すブロック図。

[図17]図１６の周期成分ピッチ波形生成部の内部を示すブロック図。

[図18]図１６の非周期成分ピッチ波形生成部の内部を示すブロック図。

[図19]図１８の非周期成分音声素片補間部の内部を示すブロック図。

[図20A]話者Ａに対応するピッチ波形の対数パワースペクトルの一例を示すグラフ図。

[図20B]図２０Ａの対数パワースペクトルの周波数を調整した場合における、話者Ａ及び話者Ｂの間のホルマントの対応関係を示す図。

[図21A]話者Ａに対応するピッチ波形の対数パワースペクトルの一例を示すグラフ図。

[図21B]図２１Ａの対数パワースペクトルのパワーを調整した場合における話者Ａ及び話者Ｂの間のホルマントの対応関係を示す図。

[図22]第６の実施形態に係る音声合成装置の最適補間比率算出部を示すブロック図。

発明を実施するための形態

[0014] 以下、図面を参照して、本発明の実施形態について説明する。

（第１の実施形態）

図１に示すように、本発明の第１の実施形態に係る音声合成装置は、有声音生成部０１、無声音生成部０２及び加算部１０１を有する。

無声音生成部０２は、音韻継続時間長００７及び音韻記号列００８に基づき無声音信号００４を生成し、加算部１０１に入力する。例えば、無声音生成部０２は、音韻記号列００８に含まれる音素が無声子音または有声摩擦音を示す場合に、当該音素に相当する無声音信号００４を生成する。無声音生成部０２の具体的構成は特に限定されないが、例えばＬＰＣ合成フィルタを白色雑音で駆動する構成を適用可能であるし、その他の既存の構成が単独で或いは組み合わされて適用されてもよい。

[0015] 有声音生成部０１は、後述するピッチマーク生成部０３、ピッチ波形生成部０４及び波形重畳部０５を有する。有声音声生成部０１には、ピッチパターン００６、音韻継続時間長００７及び音韻記号列００８が入力される。そして、有声音声生成部０１は、ピッチパターン００６、音韻継続時間長００７及び音韻記号列００８に基づき有声音信号００３を生成し、加算部１０１に入力する。

[0016] ピッチマーク生成部０３は、ピッチパターン００６及び音韻継続時間長００７に基づきピッチマーク００２を生成し、波形重畳部０５に入力する。ここで、ピッチマーク００２とは、図２に示すような、ピッチ波形００１の各々を重畳するための時間的位置を示す情報である。また、隣接するピッチマーク００２同士の間隔は、ピッチ周期に相当する。

[0017] ピッチ波形生成部０４は、ピッチパターン００６、音韻継続時間長００７及び音韻記号列００８に基づきピッチ波形００１（例えば、図２参照）を生成する。尚、ピッチ波形生成部０４の詳細な説明は後述する。

[0018] 波形重畳部０５は、ピッチマーク００２が表す時間的位置に当該ピッチマーク００２に対応するピッチ波形を重畳する（例えば、図２参照）ことにより、有声音声信号００３を生成する。波形重畳部０５は、有声音声信号００３を加算部１０１に入力する。

[0019] 加算部１０１は、有声音信号００３及び無声音信号００４を加算し、合成音声信号００５を生成し、例えばスピーカで構成される出力部（図示しない）を制御する出力制御部（図示しない）に入力する。

[0020] 以下、図３を用いてピッチ波形生成部０４を詳細に説明する。

ピッチ波形生成部０４は、最大Ｍ（Ｍは２以上の整数）人分の話者パラメータに基づく補間話者のピッチ波形００１を生成することができる。具体的には、図３に示すように、ピッチ波形生成部０４は、Ｍ個の話者パラメータ記憶部４１１，・・・，４１Ｍ、話者パラメータ選択部４２、ホルマントマッピング部４３、補間話者パラメータ４４、Ｎｌ（Ｎｌの具体的値は後述する）個の正弦波生成部４５１，…，４５Ｎｌ、Ｎｌ個の乗算部２００１，・・・２００Ｎｌ及び加算部１０２を含む。

[0021] 話者パラメータ記憶部４１ｍ（ｍは１以上Ｍ以下の任意の整数）には、話者ｍの話者パラメータが音声素片毎に分類されて記憶されている。例えば、話者ｍの音韻／ａ／に相当する音声素片の話者パラメータは、図４に示すような態様で話者パラメータ記憶部４１ｍに記憶されている。図４の例では、音韻／ａ／（その他の音韻についても同様である）に相当する音声素片は話者パラメータ記憶部４１ｍにおいて７２３１個記憶され、各音声素片には識別のための音声素片ＩＤが付与されている。第１番目の音声素片（ＩＤ＝１）は１０フレーム（ここで、１フレームはピッチ波形００１の１つ分に相当する時間的単位である。）で構成され、各フレームには識別のためのフレームＩＤが付与されている。第１番目のフレーム（ＩＤ＝１）における話者ｍの音声に相当するピッチ波形は、８個のホルマントを含んでおり、各ホルマントには識別のためのホルマントＩＤが付与されている（以降の説明において、ホルマントＩＤはホルマント周波数の昇順に増加するように付与される連続整数（初期値は「１」）とするが、ホルマントＩＤの態様はこれに限定されない。）。各ホルマントに関するパラメータとして、ホルマント周波数、ホルマント位相、ホルマントパワー及び窓関数がホルマントＩＤに対応付けられて記憶される。以降の説明において、１つのフレームを構成するホル

マントの各々のホルマント周波数、ホルマント位相、ホルマントパワー及び窓関数と、ホルマントの数とを１つのホルマントパラメータと称する。尚、各音韻に相当する音声素片の数、各音声素片を構成するフレームの数、各フレームに含まれるホルマントの数は、固定であってもよいし可変であってもよい。

[0022] 話者パラメータ選択部４２は、ピッチパターン００６、音韻継続時間長００７及び音韻記号列００８に基づき夫々１フレーム分の話者パラメータ４２１，・・・，４２Ｍを選択する。具体的には、話者パラメータ選択部４２は、話者パラメータ記憶部４１ｍに記憶されているホルマントパラメータの１つを話者ｍの話者パラメータ４２ｍとして選択して読み出す。具体的には、話者パラメータ選択部４２は、例えば図５に示すような話者ｍのホルマントパラメータを選択して話者パラメータ記憶部４１ｍから読み出す。図５の例であれば、話者パラメータ４２ｍに含まれるホルマントの数はＮｍ個であって、各ホルマントに関するパラメータとしてホルマント周波数ω、ホルマント位相Φ、ホルマントパワーａ及び窓関数ｗ（ｔ）が含まれる。話者パラメータ選択部４２は、話者パラメータ４２１，・・・，４２ｍをホルマントマッピング部４３に入力する。

[0023] ホルマントマッピング部４３は、異なる話者間でホルマントのマッピング（対応付け）を行う。具体的には、ホルマントマッピング部４３は、ある話者の話者パラメータに含まれる各ホルマントと、他の話者の話者パラメータに含まれる各ホルマントとを対応付ける。ホルマントマッピング部４３は、ホルマント同士を対応付けるためのコストを後述するコスト関数を用いて算出し、各ホルマントの対応付けを行う。但し、ホルマントマッピング部４３が行う対応付けにおいて、全てのホルマントに関して対応するホルマントが得られるとは限らない（そもそも、複数の話者パラメータ間でホルマントの数は必ずしも一致しない）。以降の説明において、ホルマントマッピング部４３は各話者パラメータにおいて夫々Ｎｌ個のホルマントの対応付けに成功するものとする。ホルマントマッピング部４３は、マッピング結果４３１を補

間話者パラメータ生成部４４に通知すると共に、話者パラメータ４２１，・・・，４２ｍを補間話者パラメータ生成部４４に入力する。

[0024] 補間話者パラメータ生成部４４は、所定の補間比率及びマッピング結果４３１に従って、補間話者パラメータを生成する。尚、補間話者パラメータ生成部４４の詳細は後述する。ここで、補間話者パラメータは、Ｎｌ個のホルマントに関するホルマント周波数４４１１，・・・，４４Ｎｌ１、ホルマント位相４４１２，・・・，４４Ｎｌ２、ホルマントパワー４４１３，・・・，４４Ｎｌ３及び窓関数４４１４，・・・，４４Ｎｌ４を含む。補間話者パラメータ生成部４４は、ホルマント周波数４４１１，・・・，４４Ｎｌ１、ホルマント位相４４１２，・・・，４４Ｎｌ２及びホルマントパワー４４１３，・・・，４４Ｎｌ３を、Ｎｌ個の正弦波生成部４５１，・・・，４５Ｎｌに夫々入力する。また、補間話者パラメータ生成部４４は、窓関数４４１４，・・・，４４Ｎｌ４をＮｌ個の乗算部２００１，・・・，２００Ｎｌに夫々入力する。

[0025] 正弦波生成部４５ｎ（ｎは１以上Ｎｌ以下の任意の整数）は、第ｎ番目のホルマントに関するホルマント周波数４４ｎ１、ホルマント位相４４ｎ２及びホルマントパワー４４ｎ３に従って正弦波４６ｎを生成する。正弦波生成部４５ｎは、正弦波４６ｎを乗算部２００ｎに入力する。乗算部２００ｎは、正弦波生成部４５ｎからの正弦波４６ｎに窓関数４４ｎ４を乗じて、第ｎ番目のホルマント波形４７ｎを得る。乗算部２００ｎは、ホルマント波形４７ｎを加算部１０２に入力する。第ｎ番目のホルマントに関するホルマント周波数４４ｎ１の値をωn、ホルマント位相４４ｎ２の値をΦn、ホルマントパワー４４ｎ３の値をａn、窓関数４４ｎ４をwn（ｔ）とし、第ｎ番目のホルマント波形４７ｎをｙn（ｔ）とすると、次の数式（１）が成立する。

[数1]

$$y_n(t) = w_n(t) \cdot a_n \cdot \cos(\omega_n t + \phi_n) \qquad (1)$$

[0026] 加算部１０２は、Ｎｌ個のホルマント波形４７１，・・・，４７Ｎｌを加算することにより、補間音声に相当するピッチ波形００１を生成する。例えば

Ｎｌの値が「３」であれば、図１１及び図１２に示すように、加算器１０２は第１番目のホルマント波形４７１と、第２番目のホルマント波形４７２と、第３番目のホルマント波形４７３とを加算することにより、補間音声に相当するピッチ波形００１を生成する。尚、図１１において点線領域で示す各グラフは、正弦波４６１，・・・，４６３、窓関数４４１４，・・・，４４３４、ホルマント波形４７１，・・・，４７３及びピッチ波形００１の時間変化（即ち、時間対振幅）を示している。また、図１２において点線領域に示す各グラフは、図１１における各グラフのパワースペクトル（即ち、周波数対振幅）を示している。このように、正弦波生成部４５１，・・・，４５Ｎｌと、乗算部２００１，・・・，２００Ｎｌと、加算部１０２とがピッチ波形合成部として作用することにより、補間音声に相当するピッチ波形００１が合成される。

[0027] 以下、ホルマントマッピング部４３が利用可能なコスト関数の一例を説明する。

ここでは、ホルマント同士を対応付けるためのコストとしてホルマント周波数及びホルマントパワーの差分に注目する。例えば、話者パラメータ選択部４２が話者Ｘの話者パラメータ４２Ｘ及び話者Ｙの話者パラメータ４２Ｙを選択したとする。話者パラメータ４２ＸにはＮx個のホルマントが含まれ、話者パラメータ４２ＹにはＮy個のホルマントが含まれている。尚、Ｎx及びＮyの値は同じでもよいし異なってもよい。このとき、話者Ｘの第ｘ番目（即ち、ホルマントＩＤ＝ｘ）のホルマントと、話者Ｙの第ｙ番目のホルマント（即ち、ホルマントＩＤ＝ｙ）のホルマントとを対応付けるコストC_{XY}（ｘ，ｙ）は、次の数式（２）で算出することができる。

[数2]

$$C_{XY}(x,y)= w_{\omega}\cdot\left({\omega_X}^{x}-{\omega_Y}^{y}\right)^2+w_a\cdot\left(\log {a_X}^{x}-\log {a_Y}^{y}\right)^2 \quad (2)$$

[0028] 数式（２）において、$\omega_X{}^x$は話者パラメータ４２Ｘに含まれる第x番目のホ

ルマントのホルマント周波数、$\omega_{Y}{}^{y}$は話者パラメータ４２Ｙに含まれる第ｙ番目のホルマントのホルマント周波数、$a_{X}{}^{x}$は話者パラメータ４２Ｘに含まれる第x番目のホルマントのホルマントパワー、$a_{Y}{}^{y}$は話者パラメータ４２Ｙに含まれる第ｙ番目のホルマントのホルマント周波数を夫々表す。また、数式（２）において、ｗωはホルマント周波数重みを表し、waはホルマントパワー重みを表す。ｗω及びwaは、設計的／実験的に導出される値を任意に設定すればよい。また、数式（２）のコスト関数はホルマント周波数の差分の２乗とホルマントパワーの差分の２乗との重み付き和であるが、ホルマントマッピング部４３が利用可能なコスト関数はこれに限られない。例えば、コスト関数は、ホルマント周波数の差分の絶対値とホルマントパワーの差分の絶対値との重み付き和であってもよいし、ホルマント同士の対応付けを評価するために有効なその他の関数を適宜組み合わせたものであってもよい。以降の説明では、特に断りのない限り、コスト関数は数式（２）を指すものとする。

[0029] 以下、図６乃至図９を用いてホルマントマッピング部４３の行うマッピング処理を説明する。ここでの説明において、ホルマントマッピング部４３は話者Ｘの話者パラメータ４２Ｘと話者Ｙの話者パラメータ４２Ｙとの間で対応付けを行うものとする。話者パラメータ４２ＸにはＮx個のホルマントが含まれ、話者パラメータ４２ＹにはＮy個のホルマントが含まれる。また、ホルマントマッピング部４３は、例えば図７に示すようなマッピング結果４３１を保持し、マッピング処理の過程においてこのマッピング結果４３１を更新する。図７に示すマッピング結果４３１において、話者Ｘの列に属する各セル（欄）には、話者パラメータ４２Ｘのホルマントの各々に対応付けられた話者パラメータ４２Ｙのホルマントのホルマント ＩＤが格納される。また、話者Ｙの列に属する各セルには、話者パラメータ４２Ｙに含まれるホルマントの各々に対応付けられた話者パラメータ４２Ｘのホルマントのホルマント ＩＤが格納される。尚、対応付けられているホルマント ＩＤが存在しない場合には、「－１」が格納される。

[0030] マッピング処理の開始時点において、いずれのホルマントも対応付けが行われていないため、マッピング結果４３１は図７に示すような状態である。マッピング処理が開始すると、ホルマントマッピング部４３は、話者パラメータ４２Ｘに含まれる全てのホルマントと、話者パラメータ４２Ｙに含まれる全てのホルマントとの間で総当たり的にコストを算出する（ステップＳ４３１）。即ち、本例であれば、ホルマントマッピング部４３は、３６組（＝９×８／２）のコストを算出する。次に、ホルマントマッピング部４３は、話者パラメータ４２Ｘに関するホルマントＩＤを指定するための変数ｘに「１」を代入し（ステップＳ４３２）、処理はステップＳ４３３に進む。

[0031] ステップＳ４３３において、ホルマントマッピング部４３は、話者パラメータ４２ＸにおけるホルマントＩＤ＝ｘのホルマントに関して、コストを最小とする話者パラメータ４２ＹのホルマントのホルマントＩＤ＝ｙminを導出する。具体的には、ホルマントマッピング部４３は次の数式（３）を計算する。

[数3]

$$y_{\min} = \arg\min_{y} C_{XY}(x, y) \qquad (3)$$

[0032] 次に、ホルマントマッピング部４３は、話者パラメータ４２ＹにおけるホルマントＩＤ＝ｙminのホルマントに関して、コストを最小とする話者パラメータ４２ＸのホルマントのホルマントＩＤ＝ｘminを導出する（ステップＳ４３４）。具体的には、ホルマントマッピング部４３は次の数式（４）を計算する。

[数4]

$$x_{\min} = \arg\min_{x'} C_{XY}(x', y_{\min}) \qquad (4)$$

[0033] 次に、ホルマントマッピング部４３は、ステップＳ４３４において導出されたｘminが変数ｘの現在の値に一致するか否かを判定する（ステップＳ４３５）。ホルマントマッピング部４３がｘmin及びｘが一致すると判定すれば処理はステップＳ４６３に進み、そうでなければ処理はステップＳ４３７に進

む。

[0034] ステップＳ４３６において、ホルマントマッピング部４３は話者パラメータ４２ＸにおけるホルマントＩＤ＝ｘ（＝ｘmin）のホルマントと話者パラメータ４２ＹにおけるホルマントＩＤ＝ｙminのホルマントとを対応付け、処理はステップＳ４３７に進む。即ち、ホルマントマッピング部４３は、マッピング結果４３１において、（行，列）＝（ｘ，話者Ｘ）で指定されるセルにｙminを格納し、（行，列）＝（ｙmin，話者Ｙ）で指定されるセルにｘを格納する。

[0035] ステップＳ４３７において、ホルマントマッピング部４３は変数ｘの現在の値がＮx未満であるか否かを判定する。ホルマントマッピング部４３が変数ｘはＮx未満であると判定すれば処理はステップＳ４３８に進み、そうでなければ処理は終了する。ステップＳ４３８において、ホルマントマッピング部４３は変数ｘを「１」インクリメントし、処理はステップＳ４３３に戻る。

[0036] ホルマントマッピング部４３によるマッピング処理の終了時点において、マッピング結果４３１は例えば図８に示すような状態である。図８に示すマッピング結果４３１において、話者パラメータ４２ＸのホルマントＩＤ＝１及び話者パラメータ４２ＹのホルマントＩＤ＝１、話者パラメータ４２ＸのホルマントＩＤ＝２及び話者パラメータ４２ＹのホルマントＩＤ＝２、話者パラメータ４２ＸのホルマントＩＤ＝４及び話者パラメータ４２ＹのホルマントＩＤ＝３、話者パラメータ４２ＸのホルマントＩＤ＝５及び話者パラメータ４２ＹのホルマントＩＤ＝４、話者パラメータ４２ＸのホルマントＩＤ＝７及び話者パラメータ４２ＹのホルマントＩＤ＝５、話者パラメータ４２ＸのホルマントＩＤ＝８及び話者パラメータ４２ＹのホルマントＩＤ＝６、話者パラメータ４２ＸのホルマントＩＤ＝９及び話者パラメータ４２ＹのホルマントＩＤ＝７が夫々対応付けられている。また、図８に示すマッピング結果４３１において、話者パラメータ４２ＸのホルマントＩＤ＝３及び８と話者パラメータ４２ＹのホルマントＩＤ＝８とで識別されるホルマントはいずれのホルマントとも対応付けられていない。

[0037] 図９において、話者パラメータ４２Ｘ及び話者パラメータ４２Ｙに対して特許文献１記載の手法を適用して得られるピッチ波形の対数パワースペクトル４３２及び４３３が夫々描かれている。対数パワースペクトル４３２及び４３３において、黒丸はホルマントを示している。そして、対数パワースペクトル４３２に含まれるホルマントの各々と対数パワースペクトル４３３に含まれるホルマントの各々とを結ぶ線が、図８に示すマッピング結果４３１に基づくホルマントの対応関係を示している。

[0038] ところで、３以上の話者パラメータに関しても、ホルマントマッピング部４３はマッピング処理を行うことができる。例えば、話者パラメータ４２Ｘ及び話者パラメータ４２Ｙに加えて更に話者Ｚに関する話者パラメータ４２Ｚをマッピング処理の対象とできる。具体的には、ホルマントマッピング部４３は、話者パラメータ４２Ｘ及び話者パラメータ４２Ｙの間と、話者パラメータ４２Ｘ及び話者パラメータ４２Ｚとの間と、話者パラメータ４２Ｙ及び話者パラメータ４２Ｚとの間で前述したマッピング処理を夫々行う。そして、話者パラメータ４２Ｘにおけるホルマント ＩＤ＝ｘと話者パラメータ４２Ｙにおけるホルマント ＩＤ＝ｙとが対応付けられ、かつ、話者パラメータ４２Ｘにおけるホルマント ＩＤ＝ｘと話者パラメータ４２Ｚにおけるホルマント ＩＤ＝ｚとが対応付けられ、かつ、話者パラメータ４２Ｙにおけるホルマント ＩＤ＝ｙと話者パラメータ４２Ｚにおけるホルマント ＩＤ＝ｚとが対応付けられていれば、ホルマントマッピング部４３はこれら３つのホルマントを互いに対応付ける。尚、マッピング処理の対象となる話者パラメータが４以上である場合にも、ホルマントマッピング部４３はマッピング処理を同様に拡張して適用すればよい。

[0039] 以下、図１０を用いて補間話者パラメータ生成部４４の行う生成処理を説明する。

補間話者パラメータ生成部４４は、話者パラメータ４２１，・・・，４２Ｍに含まれるホルマント周波数、ホルマント位相、ホルマントパワー及び窓関数を所定の補間比率を用いて補間することにより、補間話者パラメータを

生成する。ここでの説明において、補間話者パラメータ生成部４４は、話者Ｘの話者パラメータ４２Ｘと話者Ｙの話者パラメータ４２Ｙを補間比率s_X及びs_Yを夫々用いて補間するものとする。尚、補間比率s_X及びs_Yは次の数式（５）を満たす。

[数5]

$$s_X + s_Y = 1 \qquad (5)$$

[0040] 生成処理が開始すると、補間話者パラメータ生成部４４は、話者パラメータ４２Ｘに関するホルマントＩＤを指定するための変数ｘに「１」を代入し、補間話者パラメータに含まれるホルマントをカウントするための変数ＮＩに「０」を代入する（ステップＳ４４１）。そして、処理はステップＳ４４２に進む。

[0041] ステップＳ４４２において、補間話者パラメータ生成部４４は、マッピング結果４３１において話者パラメータ４２Ｘのホルマントｉｄ＝ｘに対応付けられている話者パラメータ４２Ｙのホルマントｉｄが存在するか否かを判定する。尚、図１０に示す$map_{XY}(x)$は、マッピング結果４３１において話者パラメータ４２Ｘのホルマントｉｄ＝ｘに対応付けられている話者パラメータ４２Ｙのホルマントｉｄを返す関数である。mapXY(x)が「－１」であれば、処理はステップＳ４４８に進み、そうでなければ処理はステップＳ４４３に進む。

[0042] ステップＳ４４３において、補間話者パラメータ生成部４４は、変数ＮＩを「１」インクリメントする。次に、補間話者パラメータ生成部４４は、補間話者パラメータのホルマントＩＤ（以降、便宜上補間ホルマントＩＤと称する）＝ＮＩのホルマント周波数ω_I^{NI}を算出する（ステップＳ４４４）。具体的には、補間話者パラメータ生成部４４は、次の数式（６）を計算する。

[数6]

$$\omega_I^{NI} = s_X \cdot \omega_X^{x} + s_Y \cdot \omega_Y^{mapXY(x)} \qquad (6)$$

[0043] 尚、数式（６）において、$\omega_X{}^x$は話者パラメータ４２Ｘのホルマント ＩＤ＝ｘのホルマント周波数、$\omega_Y{}^{mapXY(x)}$は話者パラメータ４２Ｙのホルマント ＩＤ＝$map_{XY}(x)$のホルマント周波数を夫々表す。

[0044] 次に、補間話者パラメータ生成部４４は、補間話者パラメータの補間ホルマント ＩＤ＝Ｎｌのホルマント位相$\Phi_l{}^{NI}$を算出する（ステップＳ４４５）。具体的には、補間話者パラメータ生成部４４は、次の数式（７）を計算する。

[数7]

$$\phi_I{}^{NI} = s_X \cdot \phi_X{}^{x} + s_Y \cdot \phi_Y{}^{mapXY(x)} \qquad (7)$$

[0045] 尚、数式（７）において、$\Phi_X{}^x$は話者パラメータ４２Ｘのホルマント ＩＤ＝ｘのホルマント位相、$\Phi_Y{}^{mapXY(x)}$は話者パラメータ４２Ｙのホルマント ＩＤ＝$map_{XY}(x)$のホルマント位相を夫々表す。

[0046] 次に、補間話者パラメータ生成部４４は、補間話者パラメータの補間ホルマント ＩＤ＝Ｎｌのホルマントパワー$a_l{}^{NI}$を算出する（ステップＳ４４６）。具体的には、補間話者パラメータ生成部４４は、次の数式（８）を計算する。

[数8]

$$a_I{}^{NI} = s_X \cdot a_X{}^{x} + s_Y \cdot a_Y{}^{mapXY(x)} \qquad (8)$$

[0047] 尚、数式（８）において、$a_X{}^x$は話者パラメータ４２Ｘのホルマント ＩＤ＝ｘのホルマントパワー、$a_Y{}^{mapXY(x)}$は話者パラメータ４２Ｙのホルマント ＩＤ＝$map_{XY}(x)$のホルマントパワーを夫々表す。

[0048] 次に、補間話者パラメータ生成部４４は補間話者パラメータの補間ホルマント ＩＤ＝Ｎｌの窓関数$w_l{}^{NI}$（ｔ）を算出し（ステップＳ４４７）、処理はステップＳ４４８に進む。具体的には、補間話者パラメータ生成部４４は、次の数式（９）を計算する。

[数9]

$$w_I^{NI} = s_X \cdot w_X^{x}(t) + s_Y \cdot w_Y^{mapXY(x)}(t) \qquad (9)$$

[0049] 尚、数式（９）において、w_X^{x}（ｔ）は話者パラメータ４２Ｘのホルマント ＩＤ＝ｘの窓関数、$w_Y^{mapXY(x)}$（ｔ）は話者パラメータ４２Ｙのホルマント ＩＤ＝$map_{XY}(x)$の窓関数を夫々表す。

[0050] ステップＳ４４８において、補間話者パラメータ生成部４４は、ｘがＮx未満であるか否かを判定する。ｘがＮx未満であれば処理はステップＳ４４９に進み、そうでなければ処理は終了する。ステップＳ４４９において、補間話者パラメータ生成部４４は変数ｘを「１」インクリメントし、処理はステップＳ４４２に戻る。尚、補間話者パラメータ生成部４４による生成処理の終了時点において、前述した変数ＮＩの値が、マッピング結果４３１において話者パラメータ４２Ｘ及び話者パラメータ４２Ｙの間で対応付けられているホルマントの数に一致していることに注意されたい。

[0051] 尚、図１０に示す生成処理は話者パラメータが３以上の場合にも拡張して適用可能である。具体的には、ステップＳ４４４乃至ステップＳ４４７において、補間話者パラメータ生成部４４は、次の数式（１０）を計算すればよい。

[数10]

$$\begin{aligned} \omega_I^{n} &= \sum_{m=1}^{M} s_m \omega_m^{map1m(x)} \\ \phi_I^{n} &= \sum_{m=1}^{M} s_m \phi_m^{map1m(x)} \\ a_I^{n} &= \sum_{m=1}^{M} s_m a_m^{map1m(x)} \\ w_I^{n}(t) &= \sum_{m=1}^{M} s_m w_m^{map1m(x)}(t) \end{aligned} \qquad (10)$$

[0052] 数式（１０）において、ｓｍは話者パラメータ４２ｍに割り当てられる補間比率をあらわしている。また、ω_I^{n}、Φ_I^{n}、a_I^{n}及びw_I^{n}（ｔ）は、補間話者パ

ラメータのホルマントＩＤ＝ｎ（１以上Ｎｌ以下の任意の整数）のホルマント周波数、ホルマント位相、ホルマントパワー及び窓関数を夫々表す。また、補間比率ｓｍは次の数式（１１）を満たすものとする。

[数11]

$$\sum_{m=1}^{M} s_m = 1 \qquad (11)$$

[0053] 以上説明したように、本実施形態に係る音声合成装置は、複数の話者パラメータ間でホルマント同士の対応付けを行い、このホルマント同士の対応関係に従って補間話者パラメータを生成している。従って、本実施形態に係る音声合成装置によれば、複数の話者パラメータ間でホルマントの位置及び数が異なる場合にも所望の声質の補間音声を合成することができる。

[0054] ここで、本実施形態に係る音声合成装置と前述した特許文献１及び特許文献２との相違点を簡潔に述べる。本実施形態に係る音声合成装置は、複数の話者パラメータに基づく補間話者パラメータを用いてピッチ波形を生成している点で特許文献１記載の音声合成方法と異なる。即ち、本実施形態に係る音声合成装置によれば、特許文献１記載の音声合成方法に比べて多くの話者パラメータを利用できるため多様な声質制御が可能となる。一方、本実施形態に係る音声合成装置は、複数の話者パラメータ間でホルマント同士の対応付けを行って、この対応関係に従って補間を行う点で特許文献２記載の音声合成装置と異なる。即ち、本実施形態に係る音声合成装置によれば、ホルマントの位置及び数が互いに異なる複数の話者パラメータを利用する場合であっても品質のよい補間音声を安定して得ることができる。

[0055] （第２の実施形態）

前述した第１の実施形態に係る音声合成装置において、補間話者パラメータ生成部４４は、ホルマントマッピング部４３による対応付けに成功したホルマントに関して補間話者パラメータを生成している。一方、本発明の第２の実施形態に係る音声合成装置における補間話者パラメータ生成部４４は、ホルマントマッピング部４３による対応付けに失敗した（即ち、他の話者パ

ラメータのいずれのホルマントとも対応付けられていない）ホルマントも上記補間話者パラメータへ挿入して利用する。

[0056] 補間話者パラメータ生成部４４による補間話者パラメータの生成処理は、図１４に示す通りである。まず、補間話者パラメータ生成部４４は、補間話者パラメータを生成（算出）する（ステップＳ４４０）。尚、ステップＳ４４０にいう補間話者パラメータは、前述した第１の実施形態と同様に、ホルマントマッピング部４３によって対応付けられているホルマントに関して生成されるものを指す。次に、補間話者パラメータ生成部４４は、各話者パラメータにおいて対応付けのされていないホルマントを、ステップＳ４４０において生成された補間話者パラメータに挿入する（ステップＳ４５０）。

[0057] 以下、図１４を用いてステップＳ４５０において補間話者パラメータ生成部４４が行う処理を説明する。

ステップＳ４５０の処理が開始すると、補間話者パラメータ生成部４４は変数mに「１」を代入し、処理はステップＳ４５２に進む（ステップＳ４５１）。ここで、変数mは、処理対象となる話者パラメータを識別する話者ＩＤを指定するための変数である。以下の説明において、話者ＩＤは、話者パラメータ記憶部４１１，・・・，４１Ｍの各々に付与される１以上Ｍ以下の互いに異なる整数とするが、これに限られるものでない。

[0058] ステップＳ４５２において、補間話者パラメータ生成部４４は変数ｎに「１」を代入し、変数ＮＵmに「０」を代入して、処理はステップＳ４５３に進む。ここで、変数ｎは、話者ＩＤ＝mの話者パラメータにおけるホルマントを識別するホルマントＩＤを指定するための変数である。また、変数ＮＵmは、図１４に示す挿入処理によって挿入された話者ＩＤ＝mの話者パラメータにおけるホルマントをカウントするための変数である。

[0059] ステップＳ４５３において、補間話者パラメータ生成部４４は、マッピング結果４３１を参照し、話者ＩＤ＝mの話者パラメータにおけるホルマントＩＤ＝ｎのホルマントが、話者ＩＤ＝１の話者パラメータにおけるいずれかのホルマントと対応付けられているか否かを判定する。具体的には、補間話

者パラメータ生成部４４は、関数$map_{1m}(n)$の返す値が「－１」であるか否かを判定する。そして、関数$map_{1m}(n)$の返す値が「－１」であれば処理はステップＳ４５４に進み、そうでなければ処理はステップＳ４５９に進む。

[0060] ステップＳ４５４において、補間話者パラメータ生成部４４は、変数ＮＵｍを「１」インクリメントする。次に、補間話者パラメータ生成部４４は、ホルマントＩＤ（以降、便宜上挿入ホルマントＩＤと称する）＝ＮＵｍのホルマント周波数ω_{Um}^{NUm}を算出する（ステップＳ４５５）。具体的には、補間話者パラメータ生成部４４は、例えば次の数式（１２）を計算する。

[数12]

$$\omega_{Um}^{NU_m} = \omega_I^{k} + \left(\omega_I^{(k+1)} - \omega_I^{k}\right) \cdot \frac{\omega_m^{n} - \omega_m^{(n-1)}}{\omega_m^{(n+1)} - \omega_m^{(n-1)}} \qquad (12)$$

[0061] 尚、数式（１２）を適用する前提として、話者ＩＤ＝ｍの話者パラメータにおけるホルマントＩＤ＝（ｎ－１）のホルマントが補間話者パラメータにおける補間ホルマントＩＤ＝ｋのホルマントの生成に用いられ、話者ＩＤ＝ｍの話者パラメータにおけるホルマントＩＤ＝（ｎ＋１）のホルマントが補間話者パラメータにおける補間ホルマントＩＤ＝（ｋ＋１）のホルマントの生成に用いられていることが必要である。数式（１２）を適用することにより、例えば図１５に示すように、話者ｍのピッチ波形の対数スペクトル４８２におけるホルマント周波数ωm^{n}と対応するように、補間話者のピッチ波形の対数スペクトル４８１におけるホルマント周波数ωUm^{NUm}が導出される。但し、このような条件を満たさない場合にも、当業者であれば数式（１２）を適宜修正して適用することにより、適切なホルマント周波数ω_{Um}^{NUm}を導出することができる。

[0062] 次に、補間話者パラメータ生成部４４は、挿入ホルマントＩＤ＝ＮＵｍのホルマント位相ϕ_{Um}^{NUm}を算出する（ステップＳ４５６）。具体的には、補間話者パラメータ生成部４４は、次の数式（１３）を計算する。

[数13]

$$\phi_{Um} = s_m \cdot \phi_m^{\,n} \qquad (13)$$

[0063] 次に、補間話者パラメータ生成部４４は、挿入ホルマントＩＤ＝ＮＵｍのホルマントパワーa_{Um}^{NUm}を算出する（ステップＳ４５７）。具体的には、補間話者パラメータ生成部４４は、次の数式（１４）を計算する。

[数14]

$$a_{Um} = s_m \cdot a_m^{\,n} \qquad (14)$$

[0064] 次に、補間話者パラメータ生成部４４は挿入ホルマントＩＤ＝ＮＵｍの窓関数a_{Um}^{NUm}を算出し（ステップＳ４５８）、処理はステップＳ４５９に進む。具体的には、補間話者パラメータ生成部４４は、次の数式（１５）を計算する。

[数15]

$$w_{Um}(t) = s_m \cdot w_m^{\,n}(t) \qquad (15)$$

[0065] ステップＳ４５９において、補間話者パラメータ生成部４４は、変数ｎの値がＮｍ未満であるか否かを判定する。変数ｎの値がＮｍ未満であれば処理はステップＳ４６０に進み、そうでなければ処理はステップＳ４６１に進む。ここで、話者ｍに関する挿入処理の終了時点において、変数ＮＵｍは次の数式（１６）満たすことに注意されたい。

[数16]

$$N_m = N_I + N_{Um} \qquad (16)$$

[0066] ステップＳ４６０において、補間話者パラメータ生成部４４は変数ｎを「１」インクリメントし、処理はステップＳ４５３に戻る。ステップＳ４６１において、補間話者パラメータ生成部４４は、変数ｍがＭ未満であるか否かを判定する。ｍがＭ未満であれば処理はステップＳ４６２に進み、そうでなければ処理は終了する。ステップＳ４６２において、補間話者パラメータ生

成部４４は変数ｍを「１」インクリメントし、処理はステップＳ４５２に戻る。

[0067] 以上説明したように、本実施形態に係る音声合成装置は、ホルマントマッピング部によって対応付けられないホルマントを補間話者パラメータに挿入している。従って、本実施形態に係る音声合成装置によれば、補間音声を合成するためにより多くのホルマントを利用できるため、補間音声のスペクトルに不連続が生じにくい。即ち、補間音声の品質を向上させることができる。

[0068] （第３の実施形態）

本発明の第３の実施形態に係る音声合成装置は、前述した第１または第２の実施形態に係る音声合成装置におけるピッチ波形生成部０４の構成を変更することにより実現される。図１６に示すように、本実施形態に係る音声合成装置におけるピッチ波形生成部０４は、周期成分ピッチ波形生成部０６、非周期成分ピッチ波形生成部０７及び加算部１０３を有する。

[0069] 周期成分ピッチ波形生成部０６は、ピッチパターン００６、音韻継続時間長００７及び音韻記号列００８に基づき、補間話者の音声の周期成分ピッチ波形０６０を生成し、加算部１０３に入力する。また、非周期成分ピッチ波形生成部０７は、ピッチパターン００６、音韻継続時間長００７及び音韻記号列００８に基づき、補間話者の音声の非周期成分ピッチ波形０７０を生成し、加算部１０３に入力する。加算部１０３は、周期成分ピッチ波形０６０及び非周期成分ピッチ波形０７０を加算してピッチ波形００１を生成し、波形重畳部０５に入力する。

[0070] 図１７に示すように、周期成分ピッチ波形生成部０６は、図３に示すピッチ波形生成部０４における話者パラメータ記憶部４１１，・・・，４１Ｍを、周期成分話者パラメータ記憶部６１１，・・・，６１Ｍに夫々置き換えて構成される。

[0071] 周期成分話者パラメータ記憶部６１１，・・・，６１Ｍには、各話者の音声に相当するピッチ波形でなく各話者の音声の周期成分に相当するピッチ波

形に関するホルマント周波数、ホルマント位相、ホルマントパワー及び窓関数などが周期成分話者パラメータとして記憶される。尚、音声を周期成分及び非周期成分に分離するための手法として、文献「P. Jackson, "Pitch-Scaled Estimation of Simultaneous Voiced and Turbulence-Noise Components in Speech," IEEE Trans. Speech and Audio Processing, vol. 9 pp. 713-726, Oct. 2001.」に記載されるものを適用可能であるが、これに限られるものでない。

[0072] 図１８に示すように、非周期成分ピッチ波形生成部０７は、非周期成分音声素片記憶部７１１，・・・，７１Ｍ、非周期成分音声素片選択部７２及び非周期成分音声素片補間部７３を有する。

[0073] 非周期成分音声素片記憶部７１１，・・・，７１Ｍには、各話者の音声の非周期成分に相当するピッチ波形（非周期成分ピッチ波形）が記憶される。

非周期成分音声素片選択部７２は、ピッチパターン００６、音韻継続時間長００７及び音韻記号列００８に基づいて、非周期成分音声素片記憶部７１１，・・・，７１Ｍに記憶されている非周期成分ピッチ波形から夫々１フレーム分の非周期成分ピッチ波形７２１，・・・，７２Ｍを選択して読み出す。非周期成分音声素片選択部７２は、非周期成分ピッチ波形７２１，・・・，７２Ｍを非周期成分音声素片補間部７３に入力する。

[0074] 非周期成分音声素片補間部７３は、非周期成分ピッチ波形７２１，・・・，７２Ｍを補間比率に従って補間し、補間話者の音声の非周期成分ピッチ波形０７０を加算部１０３に入力する。図１９に示すように、非周期成分音声素片補間部７３は、ピッチ波形接続部７４、ＬＰＣ分析部７５、パワー包絡抽出部７６、パワー包絡補間部７７、白色雑音生成部７８、乗算部２０１及び線形予測フィルタリング部７９を有する。

[0075] ピッチ波形接続部７４は、非周期成分ピッチ波形７２１，・・・，７２Ｍを時間軸方向に接続し、１つの接続済み非周期成分ピッチ波形７４０を得る。ピッチ波形接続部７４は、接続済み非周期成分ピッチ波形７４０をＬＰＣ

分析部７５に入力する。

[0076] ＬＰＣ分析部７５は、非周期成分ピッチ波形７２１，・・・，７２Ｍと接続済み非周期成分ピッチ波形７４０とに対してＬＰＣ分析を施す。そして、ＬＰＣ分析部７５は、非周期成分ピッチ波形７２１，・・・，７２Ｍの各々に対するＬＰＣ係数７５１，・・・，７５Ｍと接続済み非周期成分ピッチ波形７４０に対するＬＰＣ係数７５０とを得る。ＬＰＣ分析部７５は、ＬＰＣ係数７５０を線形予測フィルタリング部７９に入力し、ＬＰＣ係数７５１，・・・，７５Ｍをパワー包絡抽出部７６に入力する。

[0077] パワー包絡抽出部７６は、ＬＰＣ係数７５１，・・・，７５Ｍの各々に基づいてＭ個の線形予測残差波形を生成する。そして、パワー包絡抽出部７６は、線形予測残差波形の各々からパワー包絡７６１，・・・，７６Ｍを夫々抽出する。パワー包絡抽出部７６は、パワー包絡７６１，・・・，７６Ｍをパワー包絡補間部７７に入力する。

[0078] パワー包絡補間部７７は、パワー包絡７６１，・・・，７６Ｍを相関が最大となるように時間軸方向にアライメントし、これらを補間比率に従って補間することにより補間パワー包絡７７０を生成する。パワー包絡補間部７７は、補間パワー包絡７７０を乗算部２０１に入力する。

[0079] 白色雑音生成部７８は、白色雑音７８０を生成し、乗算部２０１に入力する。乗算部２０１は、白色雑音７８０に補間パワー包絡７７０を乗算する。白色雑音７８０に補間パワー包絡７７０を乗算することにより、白色雑音７８０が振幅変調され、音源波形７９０が得られる。乗算部２０１は、音源波形７９０を線形予測フィルタリング部７９に入力する。

[0080] 線形予測フィルタリング部７９は、ＬＰＣ係数７５０をフィルタ係数として用いて、音源波形７９０に対して線形予測フィルタリング処理を行って、補間話者の音声の非周期成分ピッチ波形０７０を生成する。

[0081] 以上説明したように、本実施形態に係る音声合成装置は、音声の周期成分及び非周期成分に対して異なる補間処理を施している。従って、本実施形態に係る音声合成装置によれば、前述した第１乃至第２の実施形態に比べてよ

り適切な補間が行われるため、補間音声の肉声感が向上する。

[0082] （第４の実施形態）

前述した第１乃至第３の実施形態に係る音声合成装置において、ホルマントマッピング部４３は、コスト関数として数式（２）を利用している。一方、本発明の第４の実施形態に係る音声合成装置において、ホルマントマッピング部４３は、異なるコスト関数を利用する。

[0083] 一般に、話者毎に声道長は異なり、特に、話者の性別による違いは大きい。例えば、男性音声は女性音声に比べてホルマントが低周波側に現れやすい傾向にあることが知られている。また、同性であっても、特に男性の場合は、大人の声は子供の声に比べてホルマントが低周波側に現れやすい傾向にある。このように、話者パラメータ間で声道長の差異に起因するホルマント周波数の隔たりが存在すると、マッピング処理が困難となるおそれがある。例えば、女性話者パラメータの高周波側のホルマントを男性話者パラメータの高周波側のホルマントに全く対応付けることができないおそれがある。このような場合に、例えば前述した第２の実施形態のように対応付けのされていないホルマントを補間話者パラメータとして利用したとしても、所望の声質（例えば、中性的な音声）の補間音声が得られるとは限らない。具体的には、１人の補間話者の音声でなく、２人の話者の音声であるかのような統一感のない音声が合成されてしまう。

[0084] 従って、本実施形態に係る音声合成装置において、ホルマントマッピング部４３は、次の数式（１７）をコスト関数として利用する。

[数17]

$$C_{XY}(x,y) = w_\omega \cdot \left(f\left(\omega_X{}^x\right) - \omega_Y{}^y\right)^2 + w_a \cdot \left(\log a_X{}^x - \log a_Y{}^y\right)^2 \qquad (17)$$

[0085] 数式（１７）における関数ｆ（ω）は、例えば次の数式（１８）で表される。

[数18]

$$f\left(\omega_X{}^x\right)=\alpha\cdot\omega_X{}^x \qquad (18)$$

[0086] 数式（１８）において、αは話者Ｘ及び話者Ｙの間の声道長の差異を補償する（声道長を正規化する）ための声道長正規化係数である。尚、数式（１８）において、αは、例えば話者Ｘが女性で話者Ｙが男性であれば「１」以下に設定することが望ましい。また、数式（１７）における関数ｆ（ω）は、数式（１８）に示すような線形の制御関数でなく、非線形の制御関数であってもよい。

[0087] 数式（１８）に示す関数ｆ（ω）を、図２０Ａに示す話者Ａのピッチ波形の対数パワースペクトル８０１に適用すると、図２０Ｂに示す対数パワースペクトル８０３が得られる。対数パワースペクトル８０１に関数ｆ（ω）を適用することは、対数パワースペクトル８０１を周波数軸方向に伸縮させることに相当する。このように、対数パワースペクトル８０１を周波数軸方向に伸縮させることにより話者Ａ及び話者Ｂの間の声道長の差異が補償されるため、ホルマントマッピング部４３は話者Ａの話者パラメータと話者Ｂの話者パラメータとの間でホルマントを適切にマッピングすることができる。具体的には、図２０Ｂにおいて、話者Ｂのピッチ波形の対数パワースペクトル８０２に含まれるホルマント（黒丸により図示）と対数パワースペクトル８０３に含まれるホルマント（黒丸により図示）との間を結ぶ線で表されるような対応関係を示すマッピング結果４３１が得られる。

[0088] 以上説明したように、本実施形態に係る音声合成装置は、話者間の声道長の差異を補償するようにホルマント周波数を制御したのち、ホルマントの対応付けを行う。従って、本実施形態に係る音声合成装置によれば、話者間の声道長の差異が大きい場合にも、ホルマントの対応付けが適切に行われるため、高品質な（統一感のある）補間音声を合成することができる。

[0089] （第５の実施形態）

前述した第１乃至第４の実施形態に係る音声合成装置において、ホルマン

トマッピング部４３は、コスト関数として数式（２）または数式（１７）を利用している。一方、本発明の第５の実施形態に係る音声合成装置において、ホルマントマッピング部４３は、異なるコスト関数を利用する。

[0090] 一般に、話者毎の個人差や音声の収録環境などの要因によって、話者パラメータ間で対数ホルマントパワーの平均値に差異が生じる。このように、話者パラメータ間で対数ホルマントパワーの平均値の隔たりが存在すると、マッピング処理が困難となるおそれがある。例えば、話者Ｘの話者パラメータにおける対数パワーの平均値が話者Ｙの話者パラメータにおける対数パワーの平均値に比べて小さい場合を仮定する。このとき、話者Ｘの話者パラメータにおいてホルマントパワーの比較的大きいホルマントが話者Ｙの話者パラメータにおいてホルマントパワーの比較的小さいホルマントに対応付けられる可能性がある。一方、話者Ｘの話者パラメータにおいてホルマントパワーの比較的小さいホルマント及び話者Ｙの話者パラメータにおいてホルマントパワーの比較的大きいホルマントは全く対応付けられないおそれがある。このような場合に、所望の声質（補間比率に基づき期待される声質）の補間音声が得られるとは限らない。

[0091] 従って、本実施形態に係る音声合成装置において、ホルマントマッピング部４３は次の数式（１９）をコスト関数として利用する。

[数19]

$$C_{XY}(x,y)= w_{\omega}\cdot\left(\omega_X{}^{x}-\omega_Y{}^{y}\right)^2+w_a\cdot\left(g\left(\log a_X{}^{x}\right)-\log a_Y{}^{y}\right)^2 \qquad (19)$$

[0092] 数式（１９）における関数ｇ（ｌｏｇａ）は、例えば次の数式（２０）で表される。

[数20]

$$g\left(\log a_X{}^{x}\right)=\log a_X{}^{x}+\frac{\sum \log a_Y{}^{y}}{N_Y}-\frac{\sum \log a_X{}^{x}}{N_X} \qquad (20)$$

[0093] 数式（２０）において、右辺第２項は話者Ｙの話者パラメータにおける対数ホルマントパワーの平均値、第３項は話者Ｘの話者パラメータにおける対数ホルマントパワーの平均値を夫々表す。即ち、数式（２０）は、話者Ｘ及び話者Ｙ間の対数ホルマントパワーの平均値の差異を縮小することにより話者間のパワーの差異を補償（ホルマントパワーを正規化）している。尚、数式（１９）における関数ｇ（ｌｏｇａ）は、数式（２０）に示すような線形の制御関数でなく、非線形の制御関数であってもよい。

[0094] 例えば、数式（２０）に示す関数ｇ（ｌｏｇａ）を、図２１Ａに示す話者Ａのピッチ波形の対数パワースペクトル８０１に適用すると、図２１Ｂに示す対数パワースペクトル８０４が得られる。対数パワースペクトル８０１に関数ｇ（ｌｏｇａ）を適用することは、対数パワースペクトル８０１を対数パワー軸方向に平行移動させることに相当する。このように、対数パワースペクトル８０１を対数パワー軸方向に平行移動させることにより話者Ａのパラメータ及び話者Ｂのパラメータの間における対数ホルマントパワーの平均値の差異が縮小する。故に、ホルマントマッピング部４３は話者Ａの話者パラメータと話者Ｂの話者パラメータとの間でホルマントを適切にマッピングすることができる。具体的には、図２１Ｂにおいて、対数パワースペクトル８０２に含まれるホルマントと対数パワースペクトル８０４に含まれるホルマント（黒丸により図示）との間を結ぶ線で表されるような対応関係を示すマッピング結果４３１が得られる。

[0095] 以上説明したように、本実施形態に係る音声合成装置は、話者パラメータ間で対数ホルマントパワーの平均値の差異が縮小するように対数ホルマントパワーを制御したのち、ホルマントの対応付けを行う。従って、本実施形態に係る音声合成装置によれば、話者パラメータ間で対数ホルマントパワーの平均値の差異が大きい場合にも、ホルマントの対応付けが適切に行われるため、高品質な（補間比率に基づき期待される声質に近い）補間音声を合成することができる。

[0096] （第６の実施形態）

本発明の第６の実施形態に係る音声合成装置は、前述した第１乃至第５の実施形態に従って合成する補間話者の音声を、特定の目標話者の音声に近づける最適補間比率９２１を最適補間比率算出部０９の作用によって算出する。図２２に示すように、最適補間比率算出部０９は、補間話者ピッチ波形生成部９０、目標話者ピッチ波形生成部９１及び最適補間重み算出部９２を有する。

[0097] 補間話者ピッチ波形生成部９０は、ピッチパターン００６，音韻継続時間長００７及び音韻記号列００８と、補間重みベクトル９２０により指定される補間比率とに基づき、補間音声に相当する補間話者ピッチ波形９００を生成する。補間話者ピッチ波形生成部９０の構成は、例えば図３に示すピッチ波形生成部０４と同じかこれに準ずるものでよい。但し、補間話者ピッチ波形生成部９０は、補間話者ピッチ波形９００の生成において、目標話者の話者パラメータは使用されないことに注意されたい。

[0098] ここで、補間重みベクトル９２０は、補間話者ピッチ波形生成部９０が補間話者ピッチ波形９００を生成するときに各話者パラメータに対して適用する補間比率（補間重み）を成分とするベクトルであり、例えば次の数式（２１）で表される。

[数21]

$$s = (s_1, s_2, \cdots, s_m, \cdots, s_{M-1}, s_M) \qquad (21)$$

[0099] 数式（２１）において、ｓ（左辺）は、補間重みベクトル９２０を表す。また、補間重みベクトル９２０の各成分は、次の数式（２２）を満たす。

[数22]

$$\sum_{m=1}^{M} s_m = 1 \qquad (22)$$

[0100] 目標話者ピッチ波形生成部９１は、ピッチパターン００６，音韻継続時間長００７及び音韻記号列００８と、目標話者の話者パラメータとに基づき、目標話者の音声に相当する目標話者ピッチ波形９１０を生成する。目標話者

ピッチ波形生成部９１の構成は、例えば図３に示すピッチ波形生成部０４と同じかこれに準ずるものでよいし、別の構成であってもよい。図３に示すピッチ波形生成部０４と同じ構成を利用する場合、目標話者ピッチ波形生成部９１内部の話者パラメータ選択部による話者パラメータの選択数を「１」とし、選択される話者パラメータを目標話者のものに固定すればよい（話者パラメータの選択数を特に制限せずに、目標話者に対する補間比率s_Tを「１」としてもよい）。

[0101] 最適補間重み算出部９２は、補間話者ピッチ波形９００のスペクトルと目標話者ピッチ波形９１０のスペクトルとの間の類似度を算出する。具体的には、例えば最適補間重み算出部９２は、両スペクトルの相互相関を算出する。最適補間重み算出部９２は、類似度が大きくなるように補間重みベクトル９２０をフィードバック制御する。即ち、最適補間重み算出部９２は、算出した類似度に基づき補間重みベクトル９２０を更新し、新たな補間重みベクトル９２０を補間話者ピッチ波形生成部９０に供給する。そして、最適補間重み算出部９２は、類似度が収束したときの補間重みベクトル９２０を、最適補間比率９２１として出力する。尚、類似度の収束条件は、設計的／実験的に任意に定めてよい。例えば、類似度の変動が所定の範囲内に収まったとき、類似度が所定の閾値以上に達したときなどに、最適補間重み算出部９２は類似度の収束を判定してよい。

[0102] 以上説明したように、本実施形態に係る音声合成装置は、目標話者の音声を模倣した補間音声を得るための最適補間比率を算出している。従って、本実施形態に係る音声合成装置によれば、目標話者の話者パラメータが少量であったとしてもこの目標話者の音声を模倣した補間音声を利用できるため、少量の話者パラメータから多様な声質の音声を合成することが可能となる。

[0103] 尚、本発明は上記各実施形態そのままに限定されるものではなく、実施段階ではその要旨を逸脱しない範囲で構成要素を変形して具体化できる。また上記各実施形態に開示されている複数の構成要素を適宜組み合わせることによって種々の発明を形成できる。また例えば、各実施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た構成も考えられる。さらに、異なる実施形態に記載した構成要素を適宜組み合わせてもよい。

[0104] 例えば、上記した各実施形態の処理にかかるプログラムを、コンピュータで読み取り可能な記憶媒体に格納して提供することも可能である。記憶媒体としては、磁気ディスク、光ディスク（ＣＤ－ＲＯＭ、ＣＤ－Ｒ、ＤＶＤ等）、光磁気ディスク（ＭＯ等）、半導体メモリ等、プログラムを記憶でき、且つ、コンピュータが読み取り可能な記憶媒体であれば、その記憶形式は何れの形態であってもよい。

[0105] また、上記した各実施形態の処理にかかるプログラムを、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。

[0106] その他、本発明の要旨を逸脱しない範囲で種々の変形を施しても同様に実施可能であることはいうまでもない。

符号の説明

[0107] ４２・・・話者パラメータ選択部

４２１～４２Ｍ・・・話者パラメータ

４３・・・ホルマントマッピング部

４４・・・補間話者パラメータ生成部

請求の範囲

[請求項1] 話者の音声に相当するピッチ波形毎に用意され、各ピッチ波形に含まれる複数のホルマントの各々に関するホルマント周波数、ホルマント位相、ホルマントパワー及び窓関数を含む話者パラメータを話者毎に１つずつ選択して、複数の話者パラメータを得る選択部と、

前記ホルマント周波数及び前記ホルマントパワーに基づくコスト関数を利用して前記複数の話者パラメータの間でホルマント同士の対応付けを行うマッピング部と、

前記マッピング部によって互いに対応付けられているホルマント同士でホルマント周波数、ホルマント位相、ホルマントパワー及び窓関数を所望の補間比率に従って補間して補間話者パラメータを生成する生成部と、

前記補間話者パラメータを用いて、前記補間比率に基づく補間話者の音声に相当するピッチ波形を合成する合成部と

を具備することを特徴とする音声合成装置。

[請求項2] 前記コスト関数は、前記ホルマント周波数の差分及び前記ホルマントパワーの差分の重み付き和であることを特徴とする請求項１記載の音声合成装置。

[請求項3] 前記生成部は、前記マッピング部によって対応付けられていないホルマントに関するホルマント周波数、ホルマント位相、ホルマントパワー及び窓関数を前記補間話者パラメータに挿入することを特徴とする請求項１記載の音声合成装置。

[請求項4] 前記話者パラメータは、話者の音声の周期成分に相当するピッチ波形毎に用意され、

前記合成部は、前記補間話者パラメータを用いて、前記補間話者の音声の周期成分に相当するピッチ波形を合成し、

話者の音声の非周期成分に相当する各ピッチ波形から、話者毎に１つずつ選択して複数のピッチ波形を得る第２の選択部と、

前記複数のピッチ波形を前記補間比率に従って補間して前記補間話者の音声の非周期成分に相当するピッチ波形を生成する第２の生成部と、

前記補間話者の音声の周期成分に相当するピッチ波形及び前記補間話者の音声の非周期成分に相当するピッチ波形を合成して、前記補間話者の音声に相当するピッチ波形を得る第２の合成部と

を更に具備することを特徴とする請求項１記載の音声合成装置。

[請求項5] 前記マッピング部は、話者間の声道長の差異を補償するための関数を前記ホルマント周波数に適用したうえで前記コスト関数を利用して前記複数の話者パラメータの間でホルマント同士の対応付けを行うことを特徴とする請求項１記載の音声合成装置。

[請求項6] 前記マッピング部は、話者間のパワーの差異を補償するための関数を前記ホルマントパワーに適用したうえで前記コスト関数を利用して前記複数の話者パラメータの間でホルマント同士の対応付けを行うことを特徴とする請求項１記載の音声合成装置。

[請求項7] 目標話者の音声に相当するピッチ波形を生成する第３の生成部と、

前記目標話者の音声に相当するピッチ波形に前記補間話者の音声に相当するピッチ波形を近づけるフィードバック制御を前記補間比率に対して行って、前記複数の話者パラメータに基づき前記目標話者の音声を得るための最適補間比率を算出する算出部と

を更に具備することを特徴とする請求項１記載の音声合成装置。

[請求項8] 前記補間比率は話者パラメータに割り当てられた比率であることを特徴とする請求項１記載の音声合成装置。

[請求項9] コンピュータを、

話者の音声に相当するピッチ波形毎に用意され、各ピッチ波形に含まれる複数のホルマントの各々に関するホルマント周波数、ホルマント位相、ホルマントパワー及び窓関数を含む話者パラメータを話者毎に１つずつ選択して、複数の話者パラメータを得る選択手段、

前記ホルマント周波数及び前記ホルマントパワーに基づくコスト関数を利用して前記複数の話者パラメータの間でホルマント同士の対応付けを行うマッピング手段、

前記マッピング手段によって互いに対応付けられているホルマント同士でホルマント周波数、ホルマント位相、ホルマントパワー及び窓関数を所望の補間比率に従って補間して補間話者パラメータを生成する生成手段、

前記補間話者パラメータを用いて、前記補間比率に基づく補間話者の音声に相当するピッチ波形を合成する合成手段

として機能させるための音声合成プログラム。

[請求項10] 選択部が、話者の音声に相当するピッチ波形毎に用意され、各ピッチ波形に含まれる複数のホルマントの各々に関するホルマント周波数、ホルマント位相、ホルマントパワー及び窓関数を含む話者パラメータを話者毎に１つずつ選択して、複数の話者パラメータを得ることと、

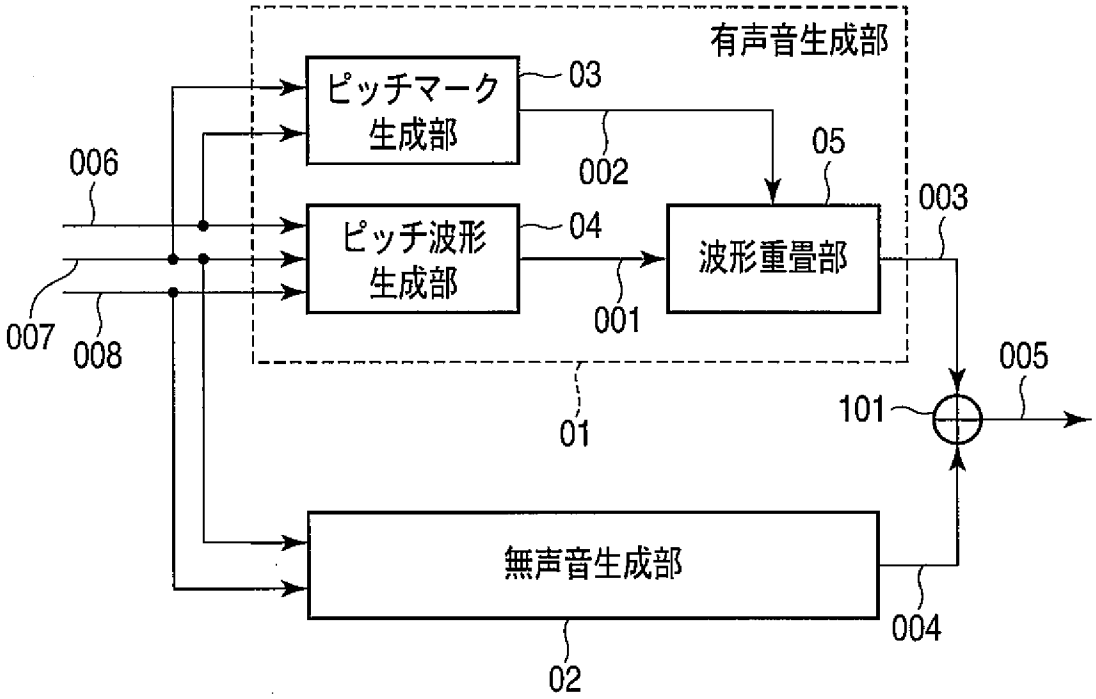
マッピング部が、前記ホルマント周波数及び前記ホルマントパワーに基づくコスト関数を利用して前記複数の話者パラメータの間でホルマント同士の対応付けを行うことと、

生成部が、前記マッピング部によって互いに対応付けられているホルマント同士でホルマント周波数、ホルマント位相、ホルマントパワー及び窓関数を所望の補間比率に従って補間して補間話者パラメータを生成することと、

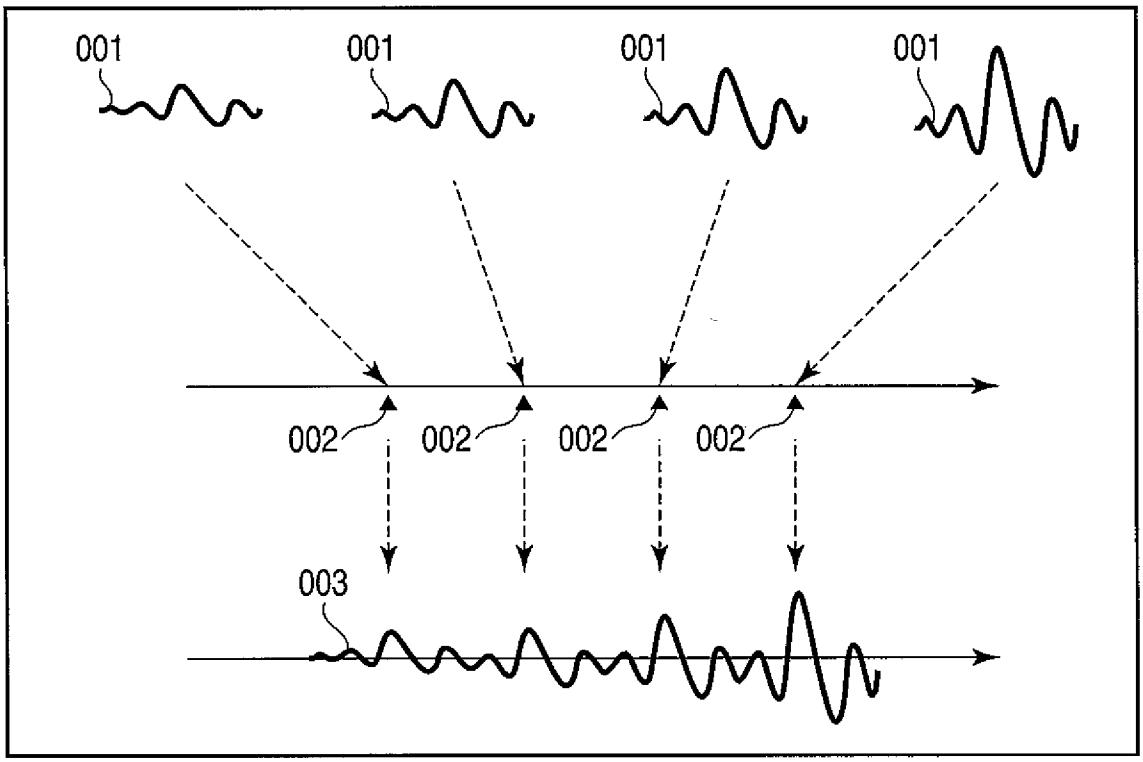
合成部が、前記補間話者パラメータを用いて、前記補間比率に基づく補間話者の音声に相当するピッチ波形を合成することと

を具備することを特徴とする音声合成方法。

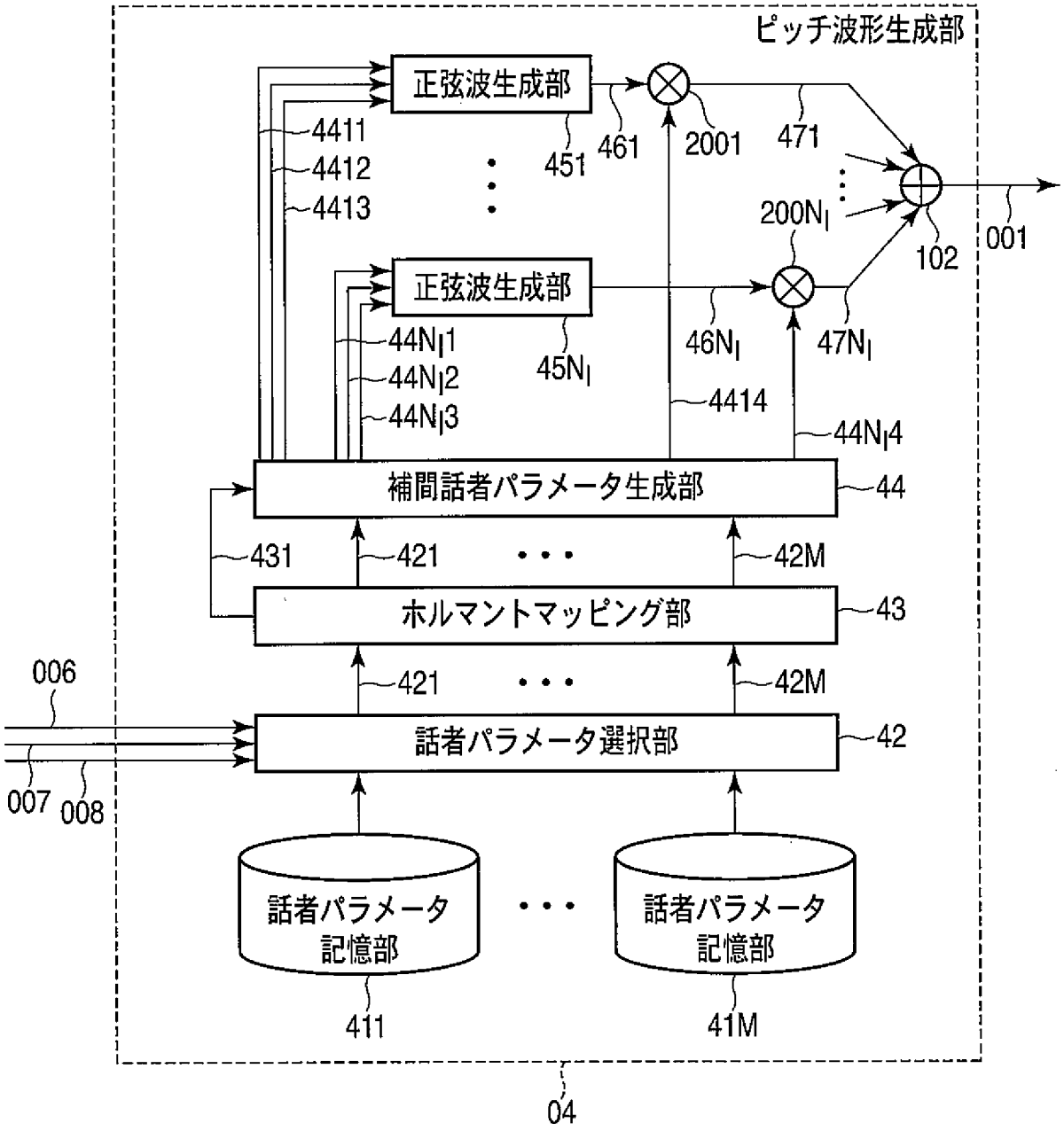
[図1]


有声音生成部
ピッチマーク生成部
03
002
05
006
ピッチ波形生成部
04
波形重畳部
003
001
007
008
005
101
01
無声音生成部
004
02

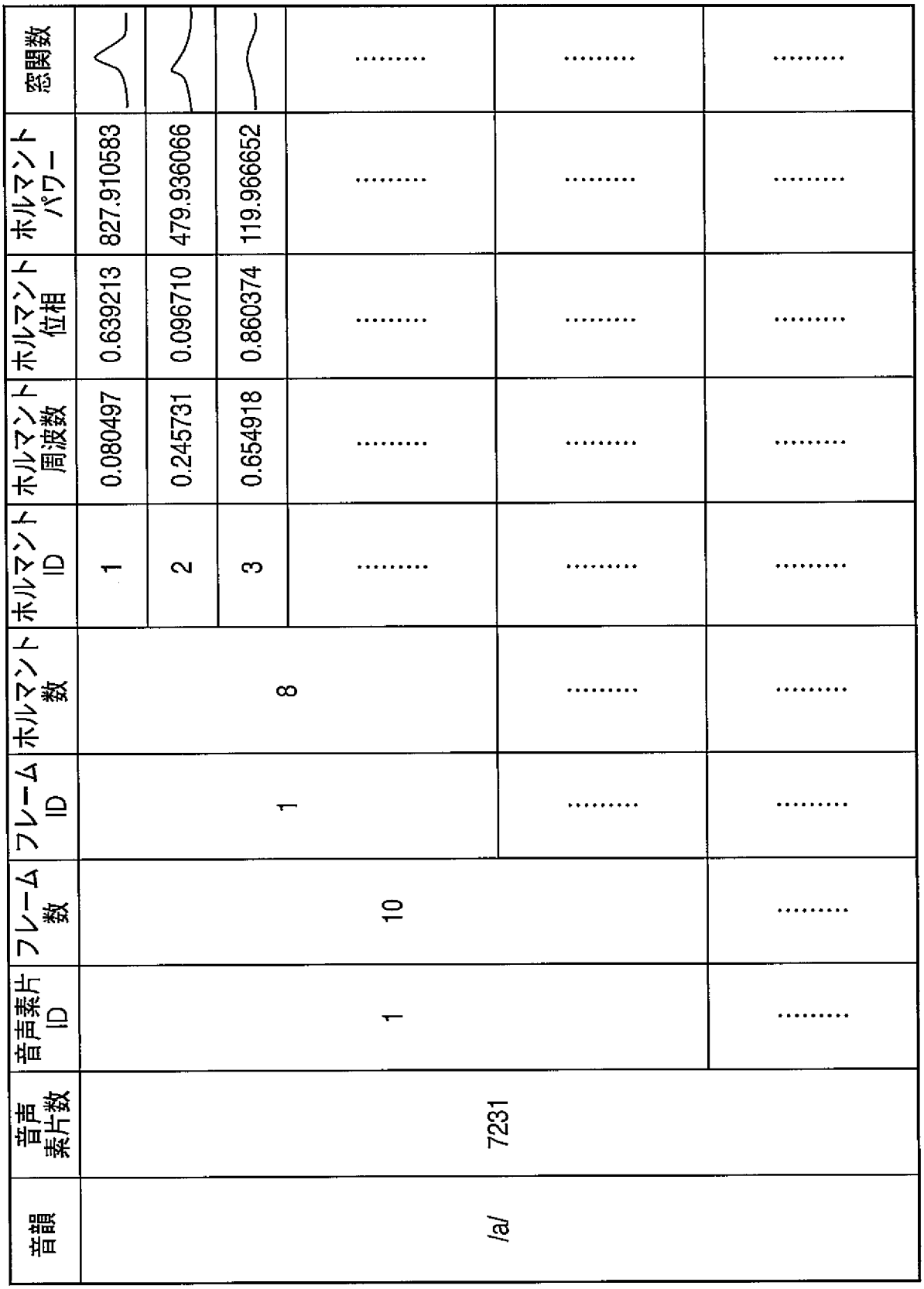
[図2]


001
001
001
001
002
002
002
002
003

[図3]

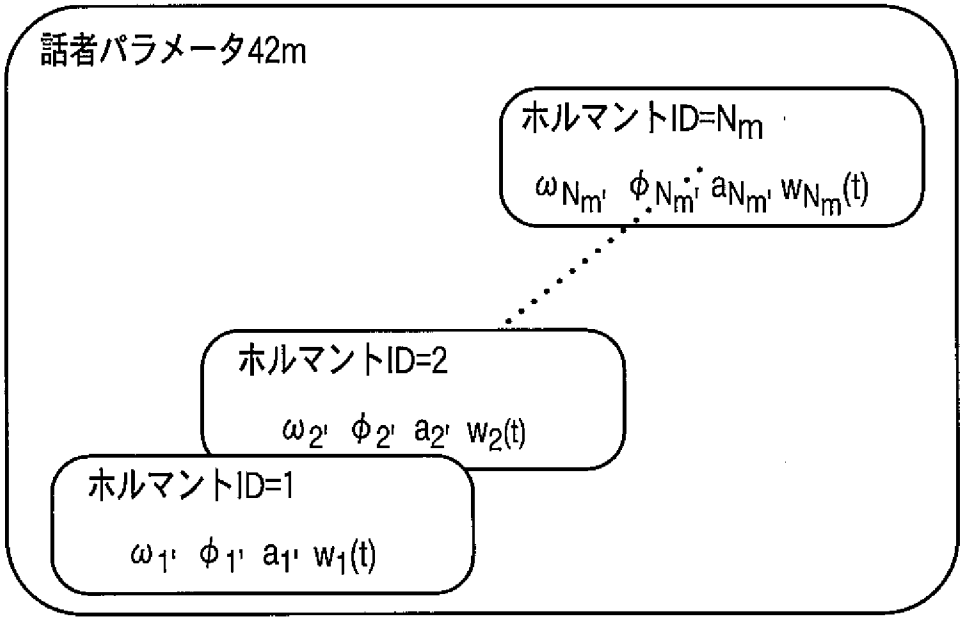

ピッチ波形生成部
正弦波生成部
4411
4412
4413
451
461
2001
471
200N_I
001
102
正弦波生成部
44N_I1
44N_I2
44N_I3
45N_I
46N_I
47N_I
4414
44N_I4
補間話者パラメータ生成部
44
431
421
42M
ホルマントマッピング部
43
006
421
42M
話者パラメータ選択部
42
007
008
話者パラメータ記憶部
話者パラメータ記憶部
411
41M
04

[図4]

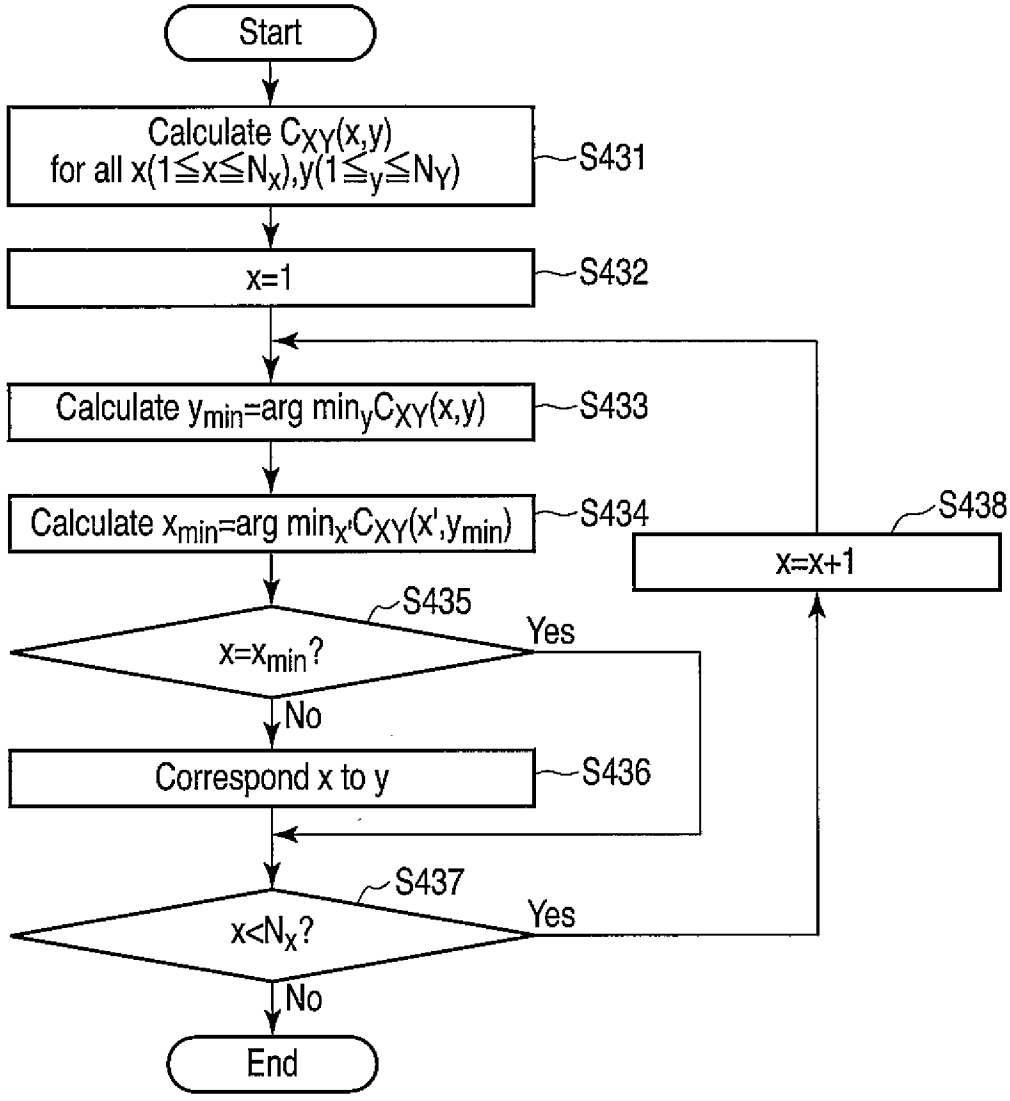


音韻	音声素片数	音声素片ID	フレーム数	フレームID	ホルマント数	ホルマントID	ホルマント周波数	ホルマント位相	ホルマントパワー	窓関数
/a/	7231	1	10	1	8	1	0.080497	0.639213	827.910583	
						2	0.245731	0.096710	479.936066	
						3	0.654918	0.860374	119.966652	
						.....	.....	.....	.....	.....
				.....	.....	.....	.....	.....	.....	.....
		.....	.....	.....	.....	.....	.....	.....	.....	.....

[図5]


話者パラメータ42m
ホルマントID=Nm
ωNm' φNm' aNm' wNm(t)
ホルマントID=2
ω2' φ2' a2' w2(t)
ホルマントID=1
ω1' φ1' a1' w1(t)

[図6]


Start
Calculate CXY(x,y) for all x(1≦x≦Nx),y(1≦y≦NY)
S431
x=1
S432
Calculate ymin=arg minyCXY(x,y)
S433
Calculate xmin=arg minx'CXY(x',ymin)
S434
S438
x=x+1
S435
x=xmin?
Yes
No
Correspond x to y
S436
S437
x<Nx?
Yes
No
End

[図7]

431

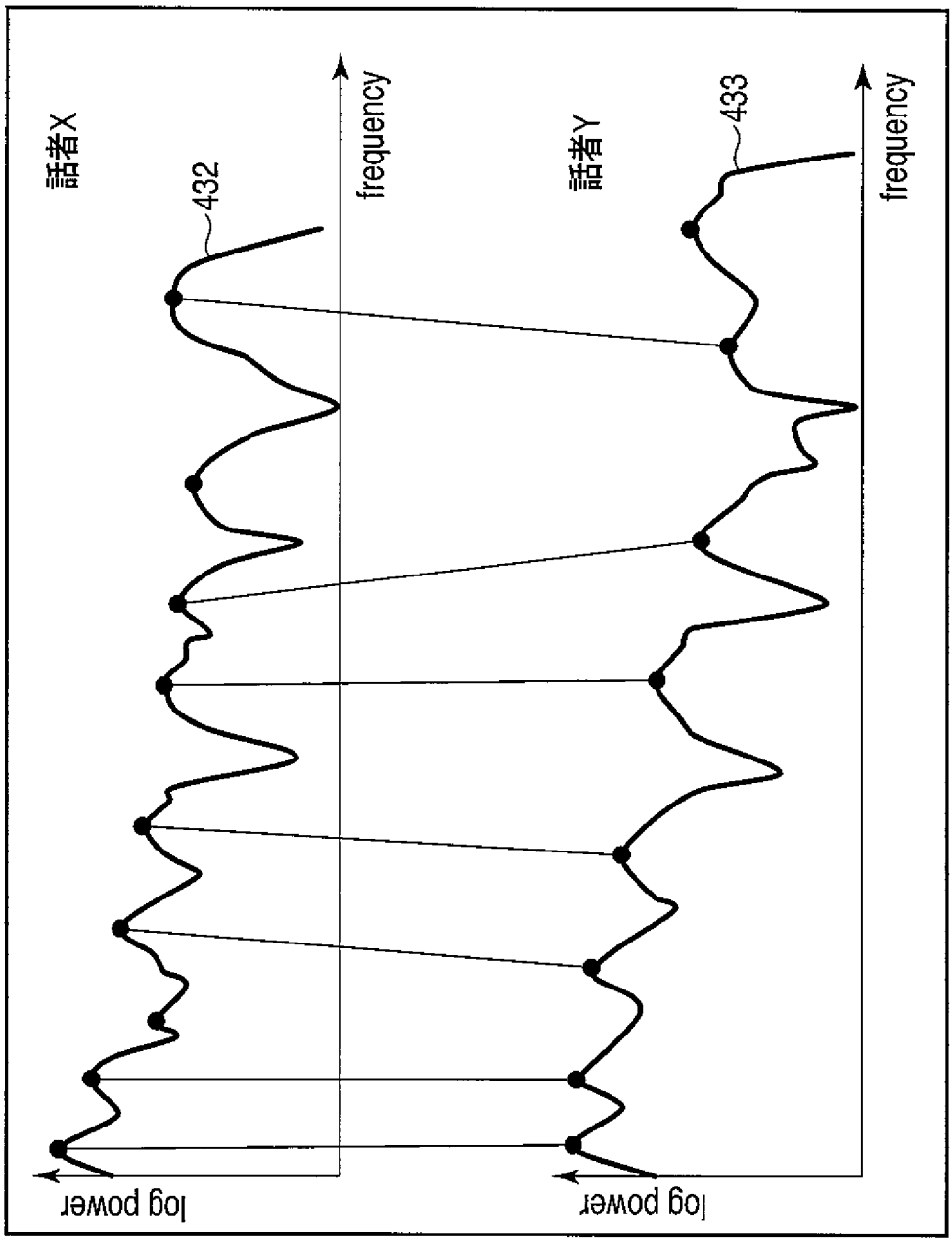
ホルマントID	話者X	話者Y
1	-1	-1
2	-1	-1
3	-1	-1
4	-1	-1
5	-1	-1
6	-1	-1
7	-1	-1
8	-1	-1
9	-1	

[図8]

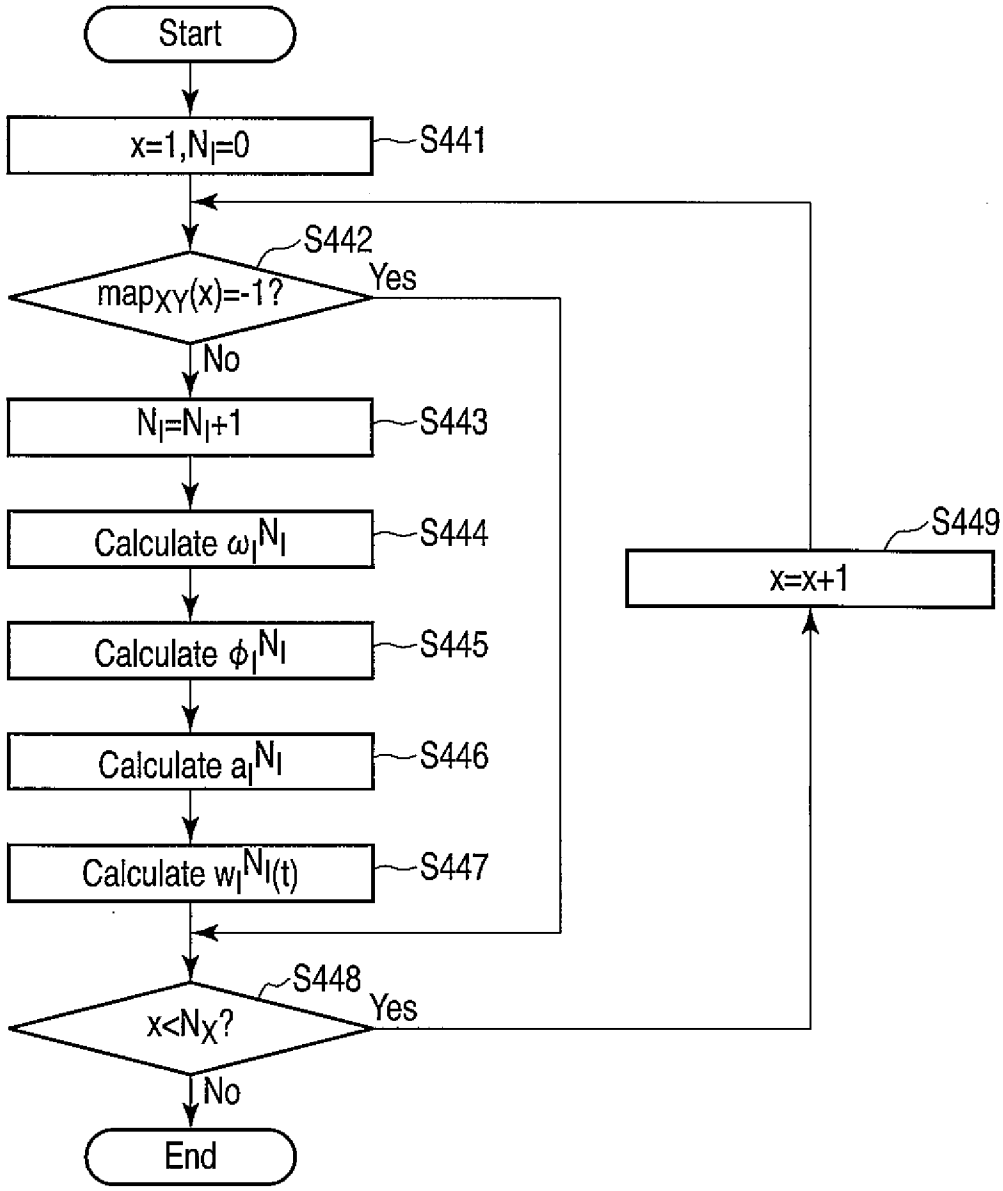
431

ホルマントID	話者X	話者Y
1	1	1
2	2	2
3	-1	4
4	3	5
5	4	6
6	5	7
7	6	9
8	-1	-1
9	7	

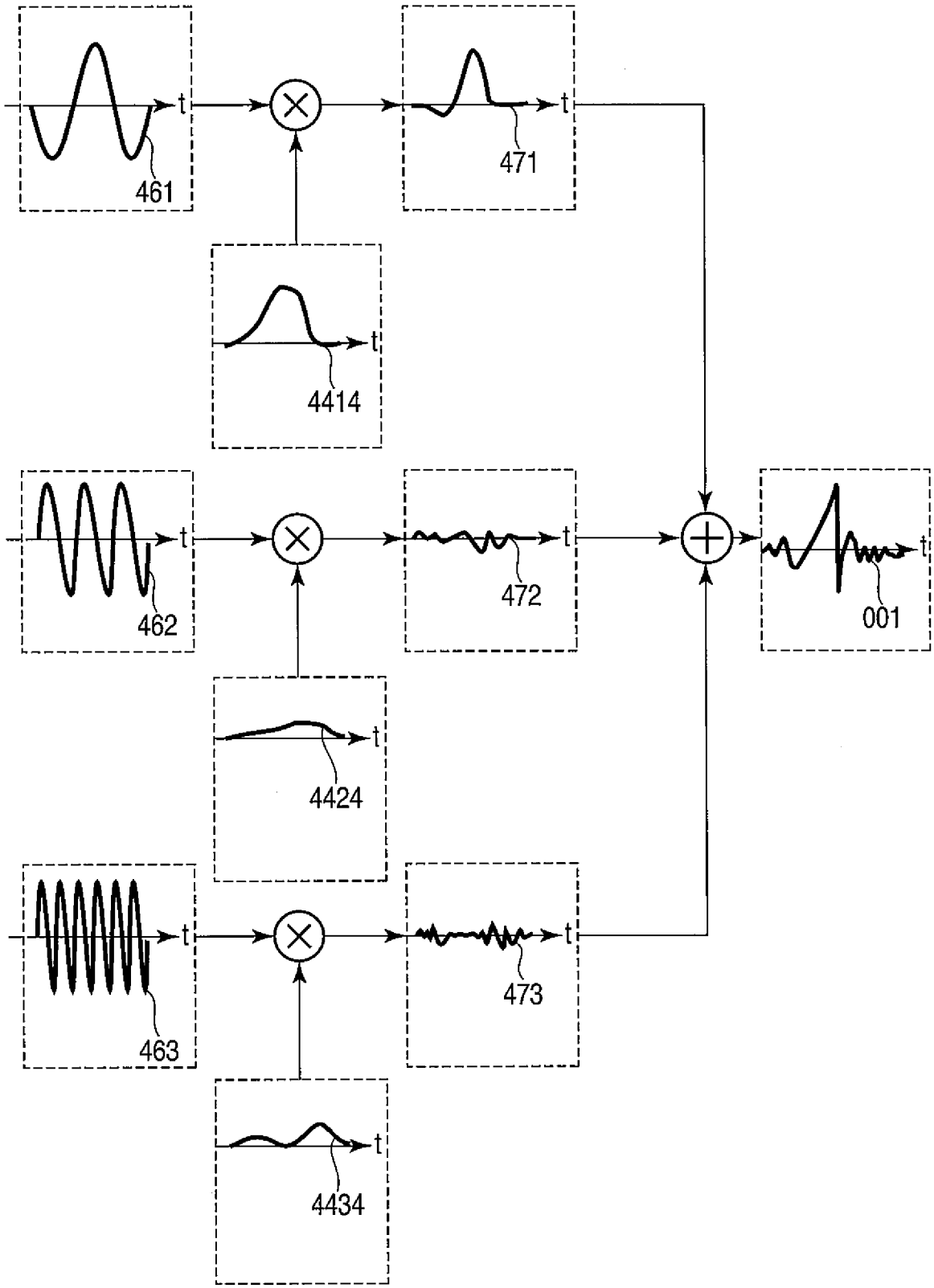
[図9]



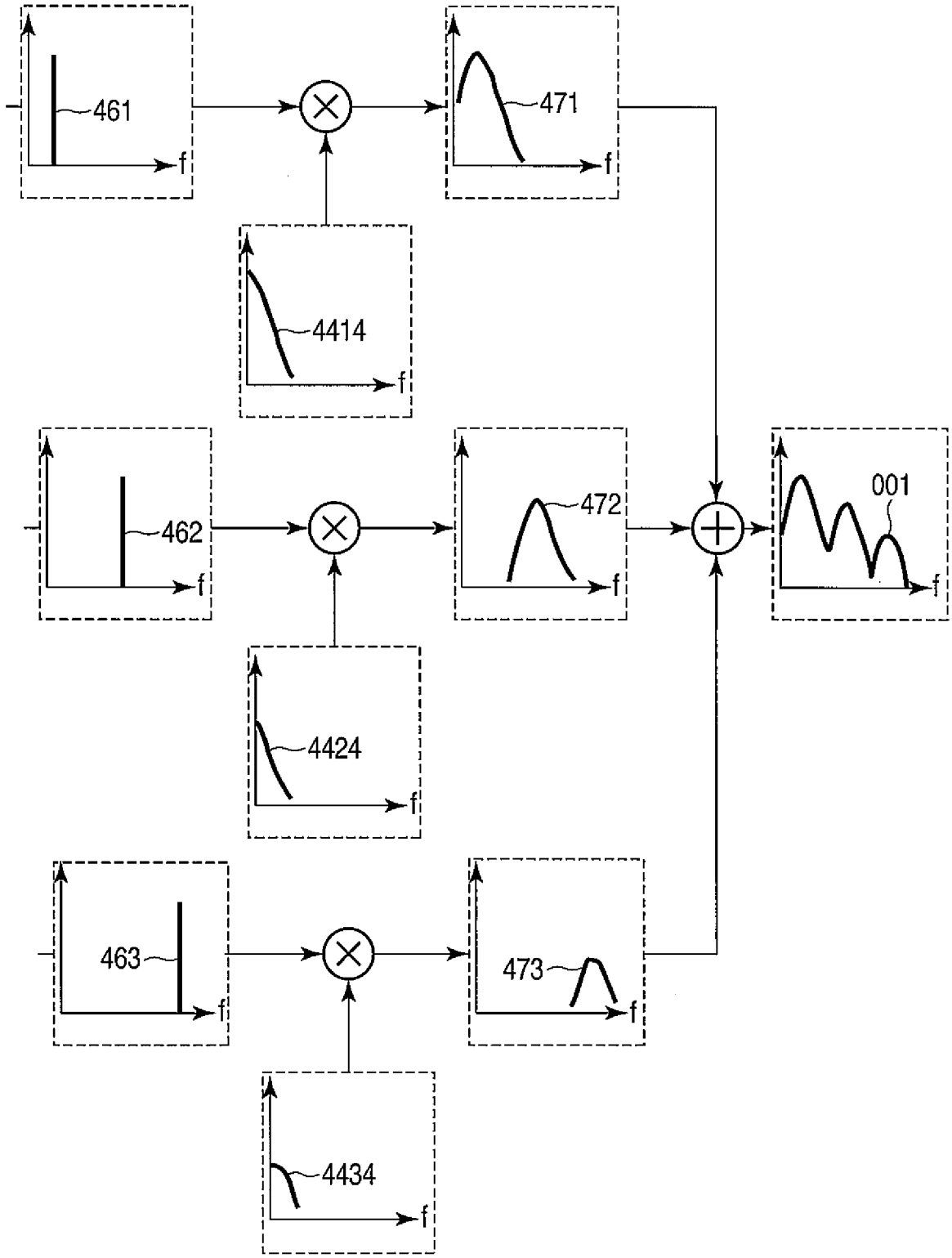
[図10]


Start
x=1,N_I=0
S441
S442
map_XY(x)=-1?
Yes
No
N_I=N_I+1
S443
Calculate ω_I^N_I
S444
Calculate φ_I^N_I
S445
Calculate a_I^N_I
S446
Calculate w_I^N_I(t)
S447
S448
x<N_X?
Yes
No
End
S449
x=x+1

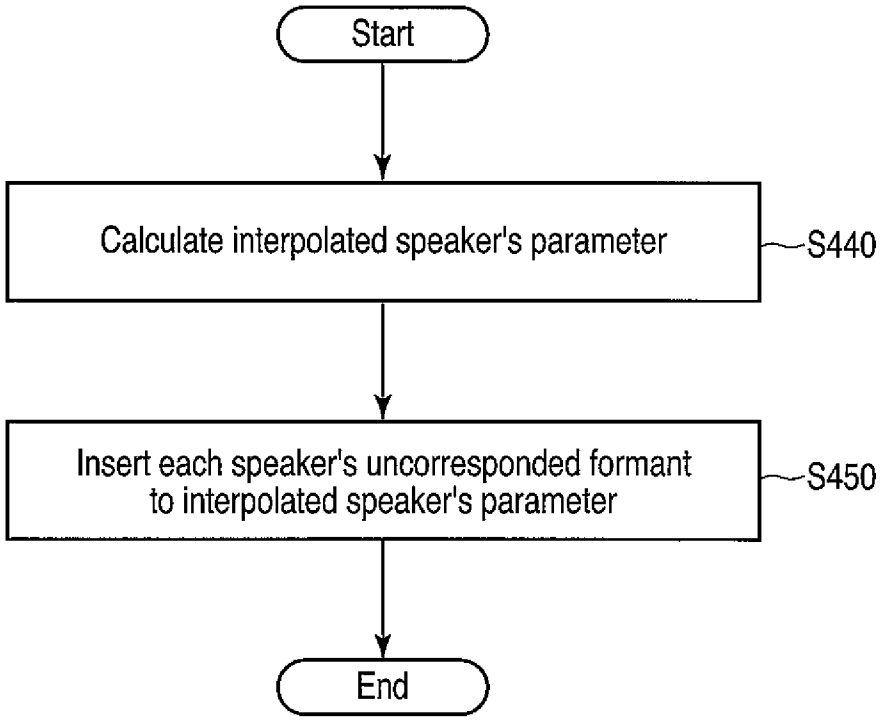
[図11]



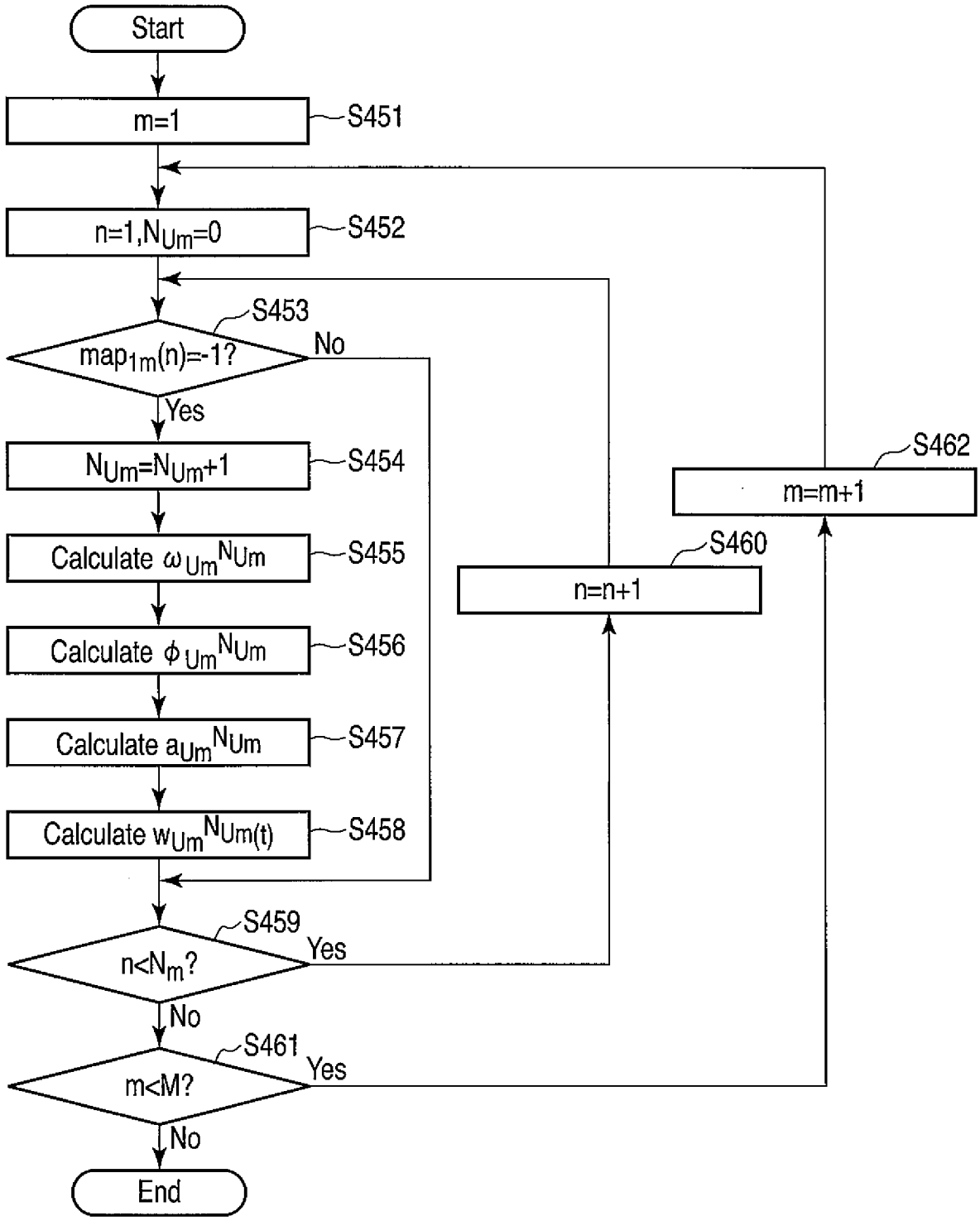
[図12]



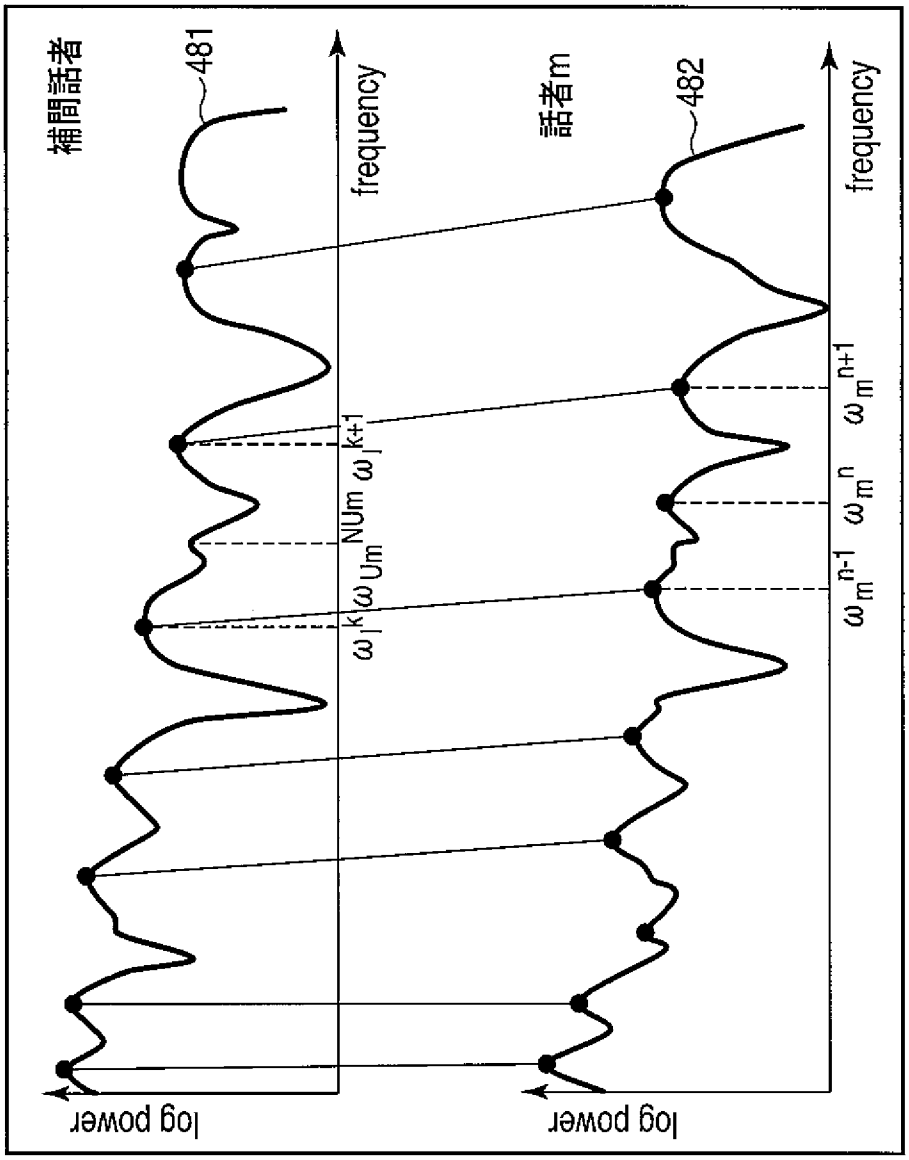
[図13]


Start
Calculate interpolated speaker's parameter
S440
Insert each speaker's uncorresponded formant to interpolated speaker's parameter
S450
End

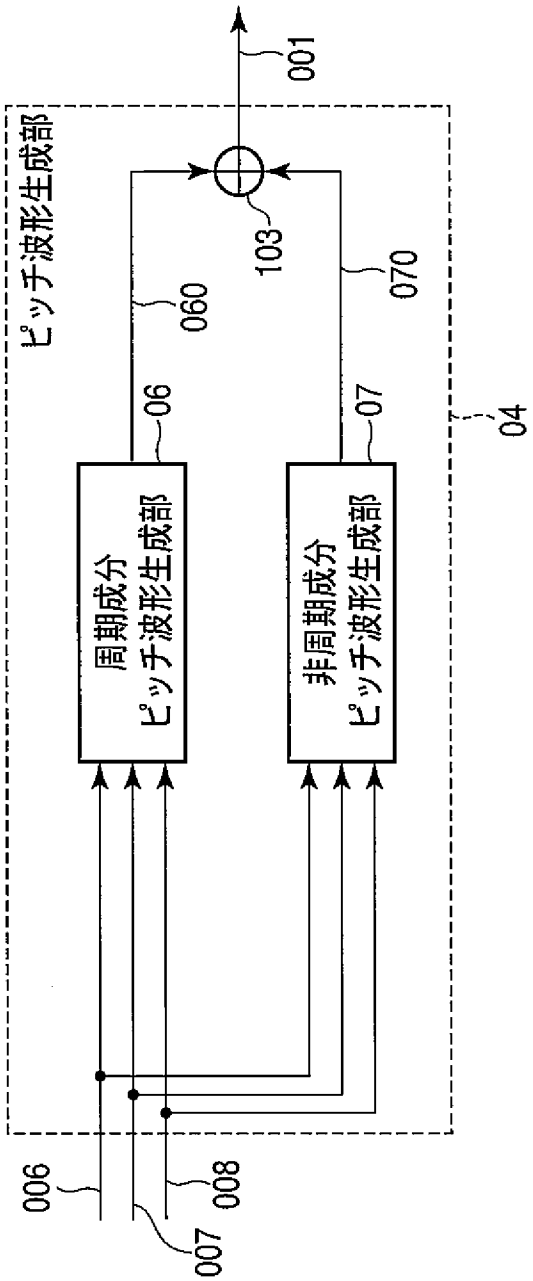
[図14]


Start
m=1
S451
n=1,N_Um=0
S452
S453
map_1m(n)=-1?
No
Yes
N_Um=N_Um+1
S454
Calculate ω_Um^N_Um
S455
Calculate φ_Um^N_Um
S456
Calculate a_Um^N_Um
S457
Calculate w_Um^N_Um(t)
S458
S459
n<N_m?
Yes
No
S461
m<M?
Yes
No
End
S460
n=n+1
S462
m=m+1

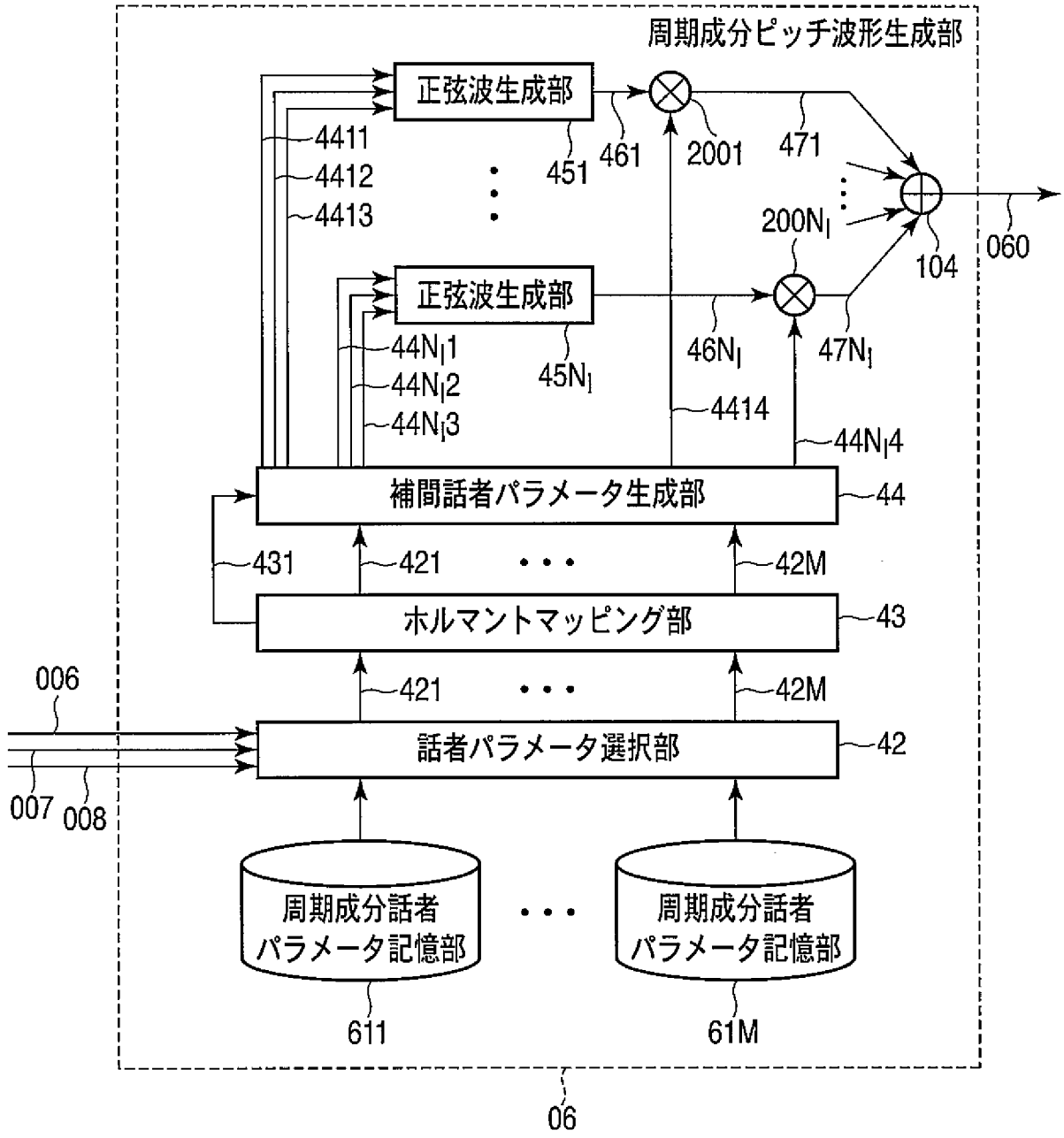
[図15]


補間話者
481
frequency
log power
話者m
482
frequency
log power
ω_j^{k} ω_{Um}^{NUm} ω_j^{k+1}
ω_m^{n-1} ω_m^{n} ω_m^{n+1}

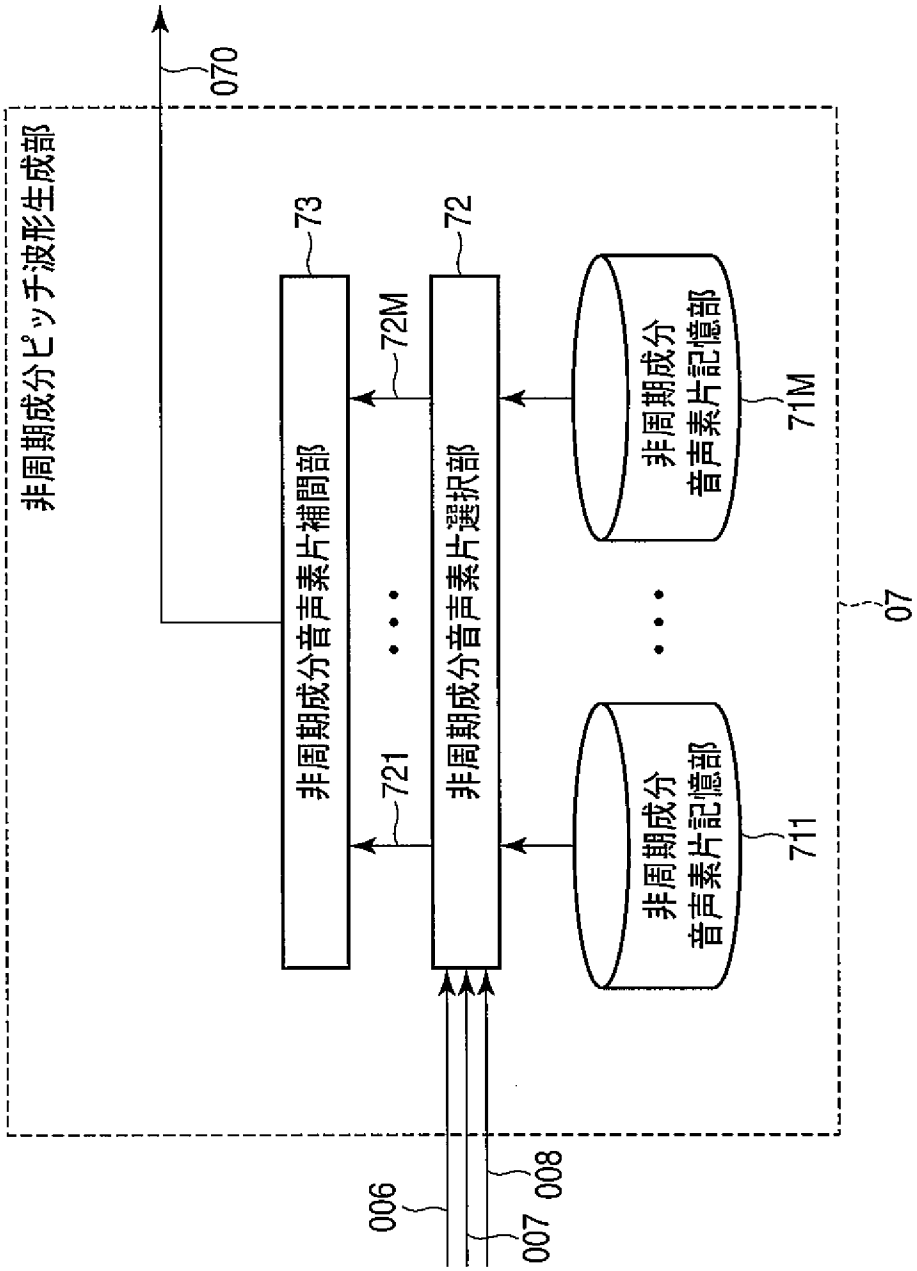
[図16]


001
ピッチ波形生成部
103
060
070
06
07
04
周期成分
ピッチ波形生成部
非周期成分
ピッチ波形生成部
006
008
007

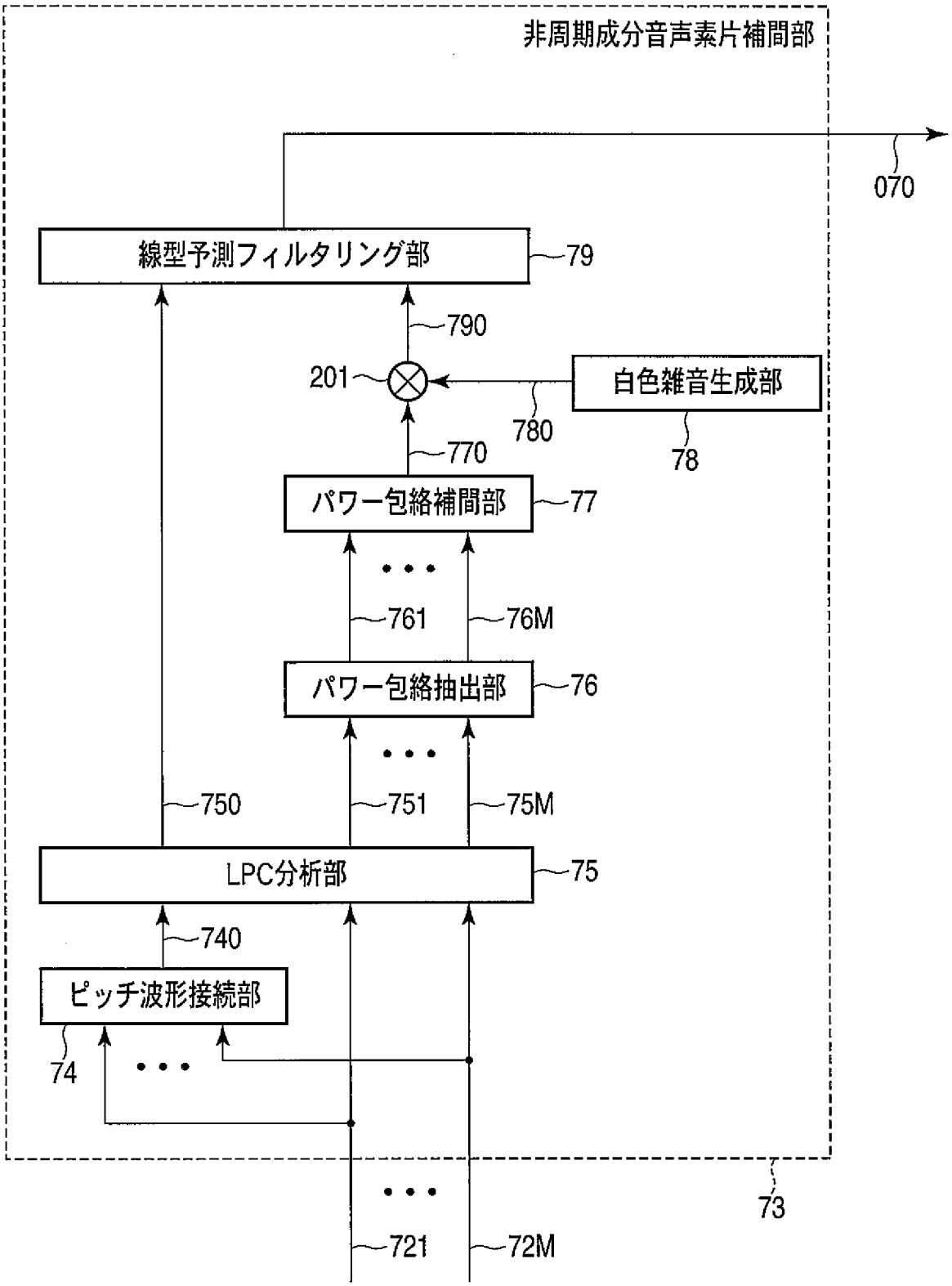
[図17]


周期成分ピッチ波形生成部
正弦波生成部
正弦波生成部
4411
4412
4413
451
461
2001
471
200N_I
104
060
44N_I1
44N_I2
44N_I3
45N_I
46N_I
47N_I
4414
44N_I4
補間話者パラメータ生成部
44
431
421
42M
ホルマントマッピング部
43
006
007
008
話者パラメータ選択部
42
周期成分話者パラメータ記憶部
周期成分話者パラメータ記憶部
611
61M
06

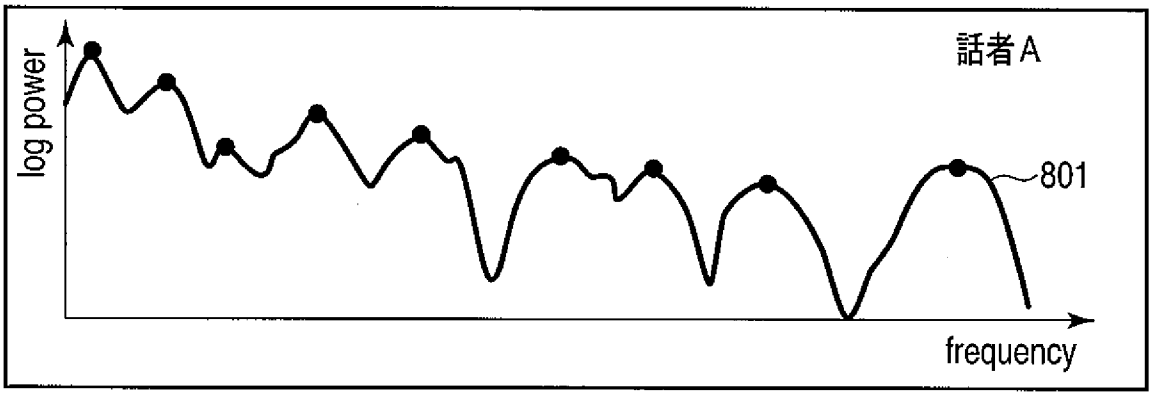
[図18]


070
非周期成分ピッチ波形生成部
73
72
72M
721
711
71M
07
非周期成分音声素片補間部
非周期成分音声素片選択部
非周期成分
音声素片記憶部
非周期成分
音声素片記憶部
006
007
008

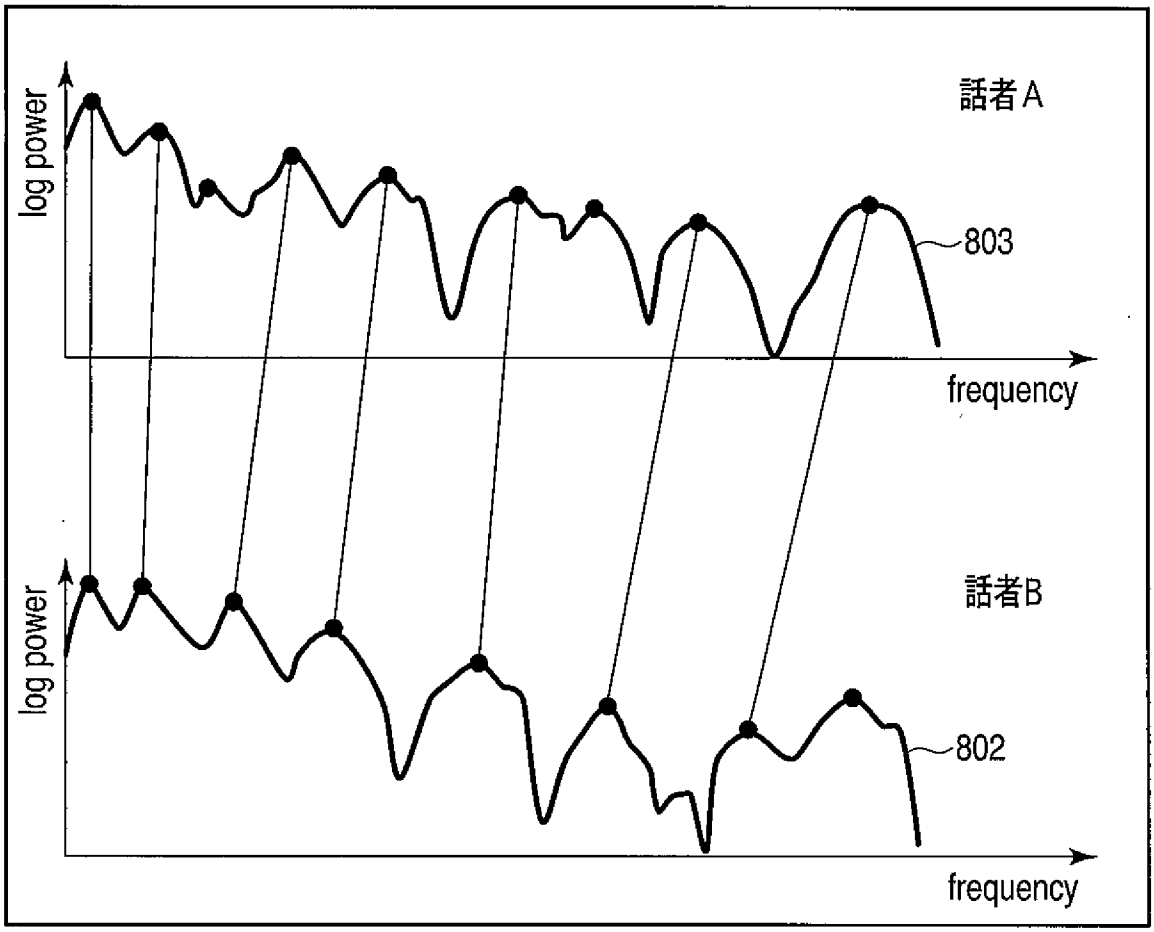
[図19]


非周期成分音声素片補間部
070
線型予測フィルタリング部
79
790
201
白色雑音生成部
780
78
770
パワー包絡補間部
77
761
76M
パワー包絡抽出部
76
750
751
75M
LPC分析部
75
740
ピッチ波形接続部
74
73
721
72M

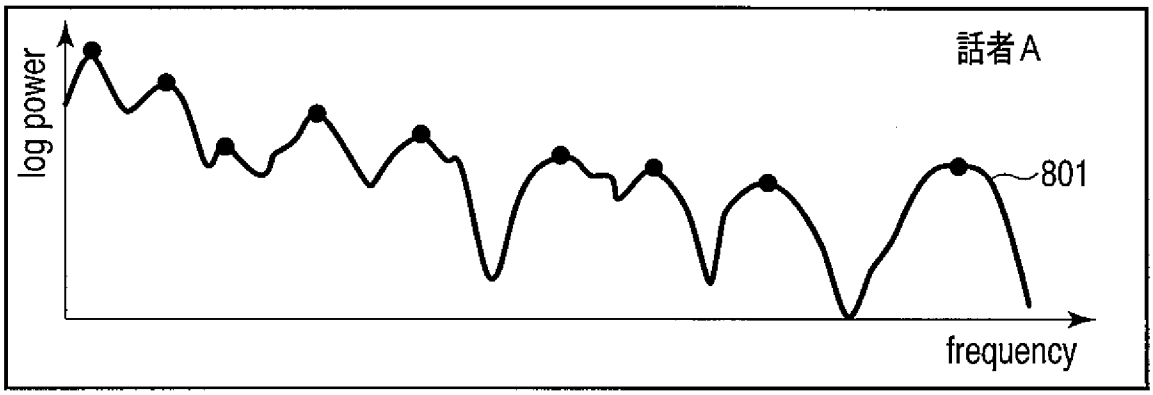
[図20A]



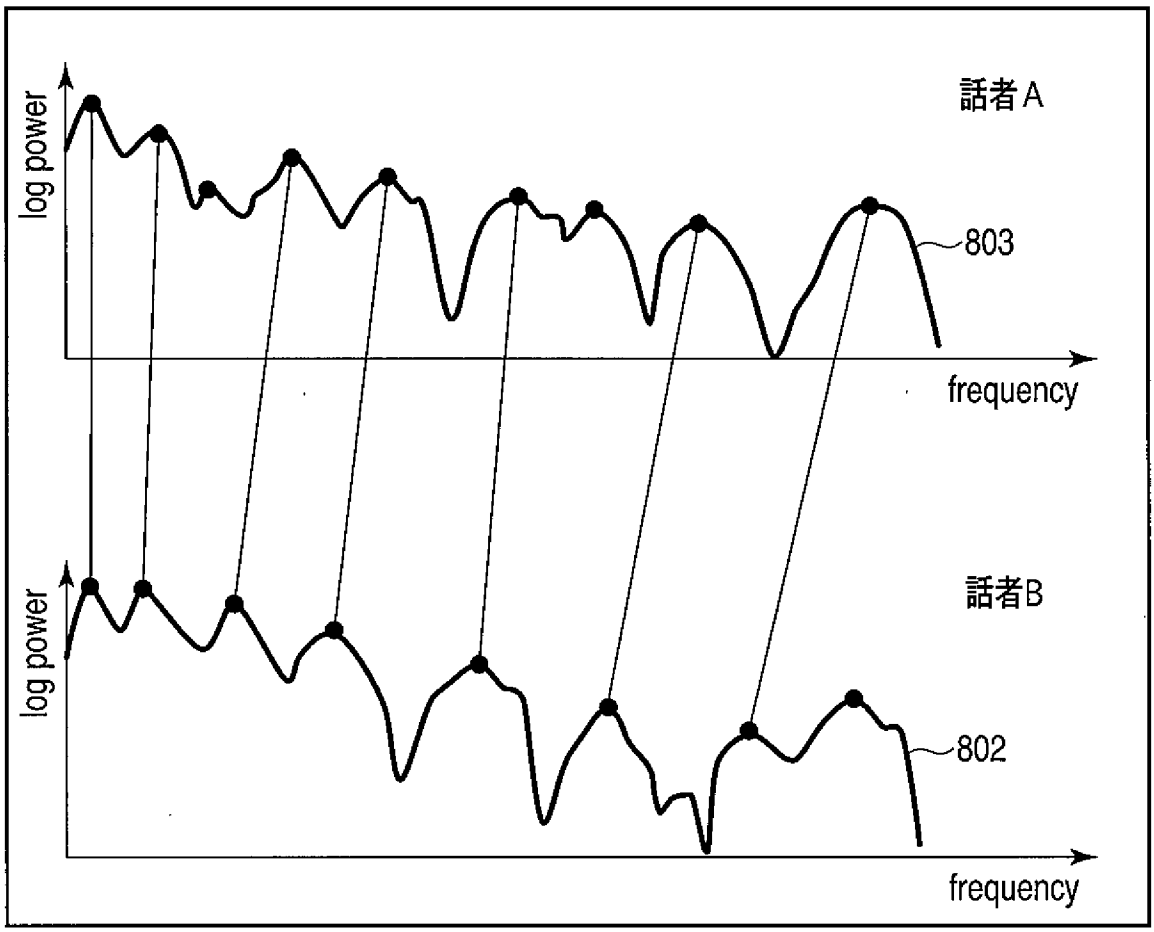
[図20B]



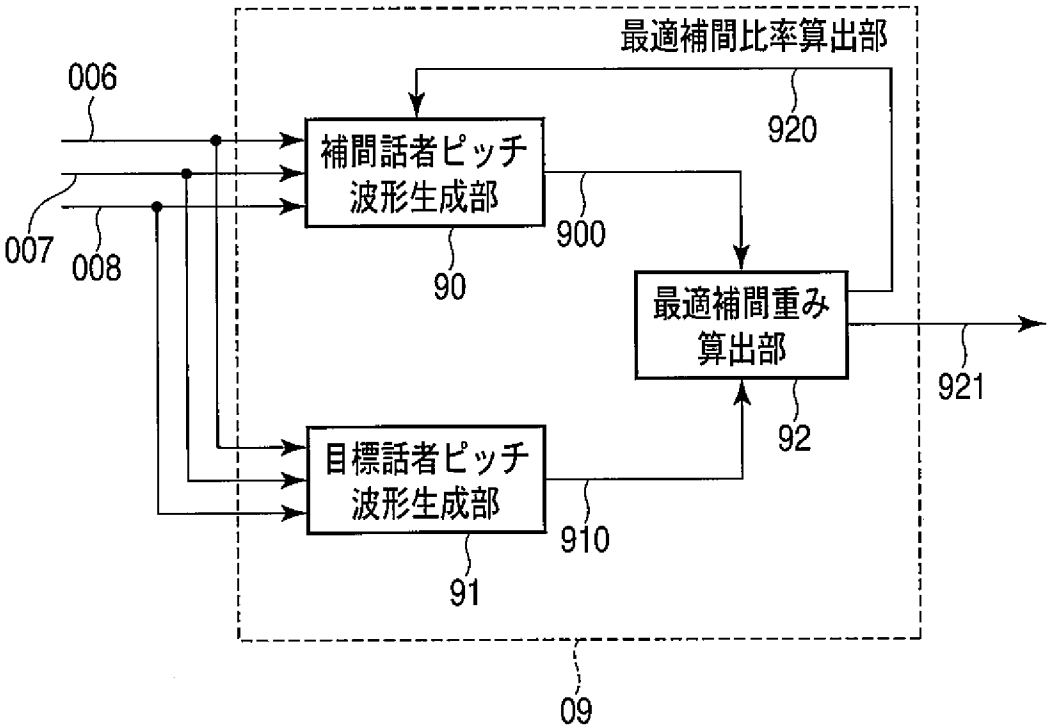
[図21A]



[図21B]


話者A
log power
803
frequency
話者B
log power
802
frequency

[図22]


最適補間比率算出部
006
007
008
補間話者ピッチ
波形生成部
90
900
920
最適補間重み
算出部
92
921
目標話者ピッチ
波形生成部
91
910
09

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2010/054250

A. CLASSIFICATION OF SUBJECT MATTER
G10L13/06(2006.01)i, *G10L21/04*(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, G10L21/04

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	1996-2010
Kokai Jitsuyo Shinan Koho	1971-2010	Toroku Jitsuyo Shinan Koho	1994-2010

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2005-43828 A (Kabushiki Kaisha Advanced Telecommunications Research Institute International), 17 February 2005 (17.02.2005), entire text; all drawings (Family: none)	1-10
A	JP 3732793 B2 (Toshiba Corp.), 21 October 2005 (21.10.2005), entire text; all drawings & JP 2002-358090 A & US 2002/0138253 A1 & EP 1246163 A2 & DE 60205421 D & DE 60205421 T & KR 10-2002-0076144 A & CN 1378199 A	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
22 April, 2010 (22.04.10)	11 May, 2010 (11.05.10)
Name and mailing address of the ISA/ Japanese Patent Office Facsimile No.	Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2010/054250

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 2951514 B2 (ATR Interpreting Telecommunications Research Laboratories), 09 July 1999 (09.07.1999), entire text; all drawings & JP 7-104791 A	1-10
A	Tatsuya MIZUTANI, "Speech synthesis based on selection and fusion of a multiple unit", THE 2004 SPRING MEETING OF THE ACOUSTICAL SOCIETY OF JAPAN Koen Ronbunshu -I-, 2004.03, pages 217 to 218	1-10
A	Ryo MORINAKA, "Speech synthesis based on the plural unit selection and fusion method using FWF model", IEICE Technical Report, 2009.01, vol.108, no.422, pages 67 to 72	1-10
P,A	JP 2009-216723 A (Kabushiki Kaisha Advanced Telecommunications Research Institute International), 24 September 2009 (24.09.2009), entire text; all drawings (Family: none)	1-10

国際調査報告

国際出願番号　ＰＣＴ／ＪＰ２０１０／０５４２５０

Ａ．　発明の属する分野の分類（国際特許分類（ＩＰＣ））

Int.Cl.　G10L13/06(2006.01)i, G10L21/04(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（ＩＰＣ））

Int.Cl.　G10L13/06, G10L21/04

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	１９７１－２０１０年
日本国実用新案登録公報	１９９６－２０１０年
日本国登録実用新案公報	１９９４－２０１０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）

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
A	JP 2005-43828 A (株式会社国際電気通信基礎技術研究所) 2005.02.17, 全文、全図（ファミリーなし）	１－１０
A	JP 3732793 B2 (株式会社東芝) 2005.10.21, 全文、全図 & JP 2002-358090 A & US 2002/0138253 A1 & EP 1246163 A2 & DE 60205421 D & DE 60205421 T & KR 10-2002-0076144 A & CN 1378199 A	１－１０
A	JP 2951514 B2 (株式会社エイ・ティ・アール音声翻訳通信研究所) 1999.07.09, 全文、全図 & JP 7-104791 A	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 ２２．０４．２０１０	国際調査報告の発送日 １１．０５．２０１０
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　９３７３ 井上　健一 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	関連する請求項の番号
Ａ	水谷　竜也　T.　Mizutani, 複数素片選択融合方式による音声合成 Speech synthesis based on selection and fusion of a multiple unit, 日本音響学会２００４年春季研究発表会講演論文集－Ｉ－ THE 2004 SPRING MEETING OF THE ACOUSTICAL SOCIETY OF JAPAN, 2004.03, ｐ．２１７－２１８	１－１０
Ａ	森中　亮　Ryo　MORINAKA, ＦＷＦモデルを用いた複数素片選択融合方式による音声合成　Speech synthesis based on the plural unit selection and fusion method using FWF model, 電子情報通信学会技術研究報告　IEICE Technical Report, 2009.01, Ｖｏｌ．１０８　Ｎｏ．４２２, ｐ．６７－７２	１－１０
Ｐ，Ａ	JP 2009-216723 A (株式会社国際電気通信基礎技術研究所) 2009.09.24, 全文、全図（ファミリーなし）	１－１０