

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2016-218281

(P2016-218281A)

(43) 公開日 平成28年12月22日 (2016. 12. 22)

(51) Int.Cl.		F I		テーマコード (参考)
<b>G 1 O L</b> 13/07 (2013.01)		G 1 O L	13/07 C	
G 1 O L 13/06 (2013.01)		G 1 O L	13/06 130	

審査請求 未請求 請求項の数 8 O L (全 16 頁)

(21) 出願番号	特願2015-103692 (P2015-103692)	(71) 出願人	000004226 日本電信電話株式会社 東京都千代田区大手町一丁目5番1号
(22) 出願日	平成27年5月21日 (2015. 5. 21)	(74) 代理人	100121706 弁理士 中尾 直樹
		(74) 代理人	100128705 弁理士 中村 幸雄
		(74) 代理人	100147773 弁理士 義村 宗洋
		(72) 発明者	水野 秀之 東京都千代田区大手町一丁目5番1号 日 本電信電話株式会社内
		(72) 発明者	井島 勇祐 東京都千代田区大手町一丁目5番1号 日 本電信電話株式会社内

(54) 【発明の名称】 音声合成装置、その方法、およびプログラム

(57) 【要約】

【課題】適切な境界周波数を自動的に決定し、合成音声の品質を向上する技術を提供する。

【解決手段】テキストに応じて素片接続される音素の素片境界前後のスペクトル特徴量の距離に基づいて境界周波数を決定し、テキストに応じた素片接続によって得られる第1合成音声のスペクトルの境界周波数に応じた高域側の成分と、テキストに音声合成のための音響モデルを適用して得られる第2合成音声のスペクトルの境界周波数に応じた低域側の成分とを混合した混合スペクトルを得る。

【選択図】 図 1

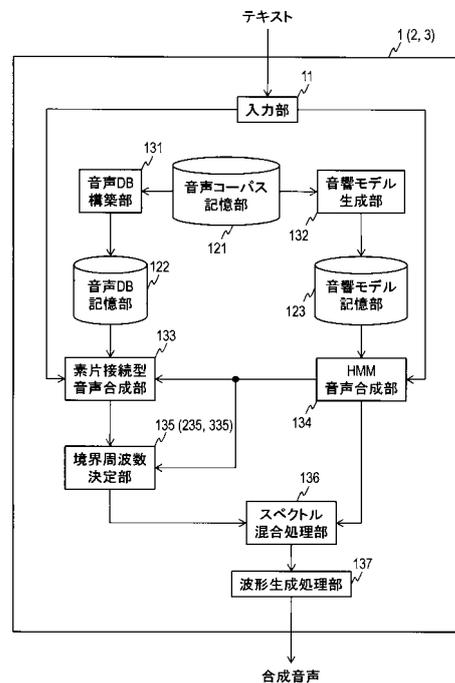


図 1

**【特許請求の範囲】****【請求項 1】**

テキストに応じて素片接続される音素の素片境界前後のスペクトル特徴量の距離に基づいて境界周波数を決定する境界周波数決定部と、

前記テキストに応じた素片接続によって得られる第 1 合成音声のスペクトルの前記境界周波数に応じた高域側の成分と、前記テキストに音声合成のための音響モデルを適用して得られる第 2 合成音声のスペクトルの前記境界周波数に応じた低域側の成分とを混合した混合スペクトルを得るスペクトル混合処理部と、  
を有する音声合成装置。

**【請求項 2】**

請求項 1 の音声合成装置であって、

前記境界周波数決定部は、

前記テキストに応じて素片接続される音素の種別に応じ、( a ) 前記スペクトル特徴量の距離に基づいて所定の周波数区間内の前記境界周波数を決定する第 1 方式、( b ) 前記周波数区間の上限値以上の周波数を前記境界周波数とする第 2 方式、または( c ) 前記周波数区間の下限値以下の周波数を前記境界周波数とする第 3 方式の何れかを選択する、音声合成装置。

**【請求項 3】**

請求項 2 の音声合成装置であって、

前記第 2 方式および前記第 3 方式は、前記テキストに応じて素片接続される音素の基本周波数に対する、前記テキストに前記音響モデルを適用して得られる基本周波数の変更割合に基づいて前記境界周波数を決定する、音声合成装置。

**【請求項 4】**

請求項 2 または 3 の音声合成装置であって、

前記第 1 方式は、前記テキストに応じて素片接続される音素のそれぞれである現音素と前記現音素の直前の先行音素との前記スペクトル特徴量の距離に基づいて前記現音素に対する前記境界周波数を決定し、

前記境界周波数決定部は、

前記現音素と前記現音素の直前の先行音素との間の歪みが音質に与える影響が大きく、かつ、前記現音素と前記現音素の直後の後続音素との間の歪みが音質に与える影響が小さい場合に、前記現音素に対して前記第 1 方式を選択し、

前記現音素と前記現音素の直後の後続音素との間の歪みが音質に与える影響が大きい場合に、前記現音素に対して前記第 2 方式を選択し、

前記現音素と前記現音素の直前の先行音素との間の歪みが音質に与える影響が小さく、かつ、前記現音素と前記現音素の直後の後続音素との間の歪みが音質に与える影響が小さい場合に、前記現音素に対して前記第 3 方式を選択する、音声合成装置。

**【請求項 5】**

請求項 1 から 4 の何れかの音声合成装置であって、

前記境界周波数決定部は、

所定の周波数区間の一部の帯域である第 1 判定帯域での前記スペクトル特徴量の距離を得、前記第 1 判定帯域での前記スペクトル特徴量の距離が前記許容限界値未満であれば、前記第 1 判定帯域に応じた周波数を前記境界周波数とし、前記第 1 判定帯域での前記スペクトル特徴量の距離が前記許容限界値未満でなければ、前記第 1 判定帯域よりも周波数の高い帯域を第 2 判定帯域とし、前記第 2 判定帯域を前記第 1 判定帯域とした処理を行う、音声合成装置。

**【請求項 6】**

請求項 5 の音声合成装置であって、

前記許容限界値は、前記第 1 判定帯域に応じた周波数に対して広義単調増加する関係にある、音声合成装置。

**【請求項 7】**

テキストに応じて素片接続される音素の素片境界前後のスペクトル特徴量の距離に基づいて境界周波数を決定する境界周波数決定ステップと、

前記テキストに応じた素片接続によって得られる第1合成音声のスペクトルの前記境界周波数に応じた高域側の成分と、前記テキストに音声合成のための音響モデルを適用して得られる第2合成音声のスペクトルの前記境界周波数に応じた低域側の成分とを混合した混合スペクトルを得るスペクトル混合処理ステップと、  
を有する音声合成方法。

【請求項8】

請求項1から6の何れかの音声合成装置としてコンピュータを機能させるためのプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は音声合成技術に関し、特に合成音声の品質を向上させる技術に関する。

【背景技術】

【0002】

近年の統計的音声合成技術の発展に伴い、高品質な合成音声の生成が可能になってきている。例えば、HMM（隠れマルコフモデル）音声合成技術（例えば非特許文献1等参照）の発展に伴い、任意の話者の音声データを学習することで、その話者の声質や調子での合成音声の生成が可能である。

【0003】

また、合成音声の高品質化技術として様々な方法（例えば非特許文献2等参照）が提案されている。しかしながら、統計的な音声合成では実際の音声の様々な現象をとらえることはできず、品質の向上には限界がある。そのため、原音声の品質を生かすことで合成音声の品質を改善する方法が提案されている。例えば非特許文献3では、原音声の品質を生かす素片接続型音声合成方式（例えば、特許文献1参照）とHMM音声合成方式を組み合わせ、高域部分に素片接続型音声合成方式によるスペクトルを利用し、低域部分にHMM音声合成方式によるスペクトルを利用することで、合成音声の品質を改善する方法が提案されている。

【先行技術文献】

【特許文献】

【0004】

【特許文献1】特許第2761552号公報

【非特許文献】

【0005】

【非特許文献1】H. Zen, K. Tokuda, T. Masuko, T. Kobayashi and T. Kitamura, "A Hidden semi-Markov model-based speech synthesis system," IEICE Trans. Inf. and Syst., vol.E90-D, no.5, pp.825-834, 2007.

【非特許文献2】T. Toda., and K.Tokuda. "A speech parameter generation algorithm considering global variance for HMM-based speech synthesis." IEICE Trans. Inf. and Syst vol.E90-D, no.5, pp.816-824, 2007.

【非特許文献3】T. Inoue, S.Hara, M.Abe, "A Hybrid Text-to-Speech Based on Sub-Band Approach," in Proc of APSIPA DOI:10.1109/APSIPA.2014.7041575, 2014.

【発明の概要】

【発明が解決しようとする課題】

【0006】

非特許文献3の方法では、高域部分と低域部分の接続境界となる境界周波数が合成音声の品質に大きな影響を与える。すなわち、境界周波数を下げるほど、合成音声に占める素片接続型音声合成方式によるスペクトルが増加し、HMM音声合成方式によるスペクトルが減少する。そのため、境界周波数を下げるほど、合成音声の自然感（肉声感）は向上す

10

20

30

40

50

るが、隣接する音素（素片）の接続部での歪みに基づく異音発生頻度が高くなる。一方、境界周波数を上げるほど、合成音声に占める素片接続型音声合成方式によるスペクトルが減少し、HMM音声合成方式によるスペクトルが増加する。そのため、境界周波数を上げるほど、音素の接続部での歪みに基づく異音発生頻度は低くなるが、合成音声の自然感は低下する。よって、音素の接続部での歪みに基づく異音発生と、素片接続型音声合成方式導入による自然感の向上との両方を考慮しなければ、合成音声の品質を向上することはできない。

【0007】

これらを考慮した適切な境界周波数は素片接続される音素に依存する。しかしながら、非特許文献3の方法では一律の境界周波数が設定されていたため、十分な品質向上効果が得られない場合があった。また、従来は境界周波数を手動で設定する必要があった。

10

【0008】

本発明の課題は、適切な境界周波数を自動的に決定し、合成音声の品質を向上する技術を提供することである。

【課題を解決するための手段】

【0009】

テキストに応じて素片接続される音素の素片境界前後のスペクトル特徴量の距離に基づいて境界周波数を決定し、テキストに応じた素片接続によって得られる第1合成音声のスペクトルの境界周波数に応じた高域側の成分と、テキストに音声合成のための音響モデルを適用して得られる第2合成音声のスペクトルの境界周波数に応じた低域側の成分とを混合した混合スペクトルを得る。

20

【発明の効果】

【0010】

これにより、適切な境界周波数を自動的に決定し、合成音声の品質を向上できる。

【図面の簡単な説明】

【0011】

【図1】図1は実施形態の音声合成装置を例示したブロック図である。

【図2】図2Aは実施形態の境界周波数決定部を例示したブロック図である。図2Bは実施形態の波形生成処理部を例示したブロック図である。

【図3】図3は実施形態の音声合成方法を例示するための図である。

30

【図4】図4Aは実施形態の境界周波数決定方法を例示するためのフロー図である。図4Bは実施形態の境界周波数決定方法を例示するための概念図である。

【図5】図5は実施形態の波形生成処理を例示するための概念図である。

【図6】図6は実施形態の境界周波数決定方法を例示するためのフロー図である。

【図7】図7Aおよび図7Bは実施形態の境界周波数決定方法を例示するための概念図である。

【発明を実施するための形態】

【0012】

以下、本発明の実施形態を説明する。

[概要]

40

実施形態の概要を説明する。実施形態では、入力された「テキスト」に応じて素片接続される音素の素片境界前後のスペクトル特徴量の距離に基づいて境界周波数を決定し、「テキスト」に応じた素片接続によって得られる第1合成音声のスペクトルの境界周波数に応じた高域側の成分と、「テキスト」に音声合成のための「音響モデル」を適用して得られる第2合成音声のスペクトルの境界周波数に応じた低域側の成分とを混合した混合スペクトルを得る。「音響モデル」の例はHMMなどの確率モデルである。境界周波数よりも高域の帯域が「境界周波数に応じた高域側」であり、それ以外の帯域が「境界周波数に応じた低域側」であってもよいし、境界周波数以上の高域が「境界周波数に応じた高域側」であり、それ以外の帯域が「境界周波数に応じた低域側」であってもよい。境界周波数よりも高域の帯域が「境界周波数に応じた高域側」であり、境界周波数よりも低域の帯域が

50

「境界周波数に応じた低域側」であってもよい。境界周波数に定数または変数を加算または減算した周波数を境界として「境界周波数に応じた高域側」および「境界周波数に応じた低域側」が定められてもよい。「境界周波数に応じた高域側」の帯域と「境界周波数に応じた低域側」の帯域とが一部で重複してもよい。ここで、素片境界前後のスペクトル特徴量の距離に基づいて境界周波数を決定するため、境界周波数の設定を自動化できるとともに、音素の接続部での歪の大きさに応じ、合成音声に占める素片接続型音声合成方式によるスペクトルが含まれる帯域を調整できる。その結果、音素の接続部での歪に基づく異音発生と、素片接続型音声合成方式導入による自然感の向上との両方を考慮して合成音声の品質を向上させることができる。

【0013】

「テキスト」に応じて素片接続される音素の種別に応じ、(a)スペクトル特徴量の距離に基づいて「所定の周波数区間」内の境界周波数を決定する第1方式、(b)「所定の周波数区間」の上限値以上の周波数を境界周波数とする第2方式、または(c)「所定の周波数区間」の下限値以下の周波数を境界周波数とする第3方式の何れかを選択してもよい。これにより、素片接続される音素の種別に応じて適切な境界周波数を選択でき、合成音声の品質を向上させることができる。

【0014】

例えば、「テキスト」に応じて素片接続される音素のそれぞれである「現音素」と「現音素」の直前の「先行音素」との間の歪みが音質に与える影響が大きく、かつ、「現音素」と「現音素」の直後の「後続音素」との間の歪みが音質に与える影響が小さい場合、「現音素」に対する境界周波数を決定するために「第1方式」が選択される。例えば「現音素」が母音または有声子音を表し、「現音素」の直前の「先行音素」が母音または有声子音を表し、「現音素」の直後の「後続音素」が無声子音を表す場合に、「現音素」に対して「第1方式」が選択される。これらの例の「第1方式」では「現音素」と「現音素」の直前の「先行音素」とのスペクトル特徴量の距離に基づいて「現音素」に対する境界周波数を決定する。これにより、「現音素」と「先行音素」との間の歪みに基づく異音発生と、素片接続型音声合成方式導入による自然感の向上との両方を考慮し、「現音素」に対応する合成音声の品質を向上できる。

【0015】

例えば、「現音素」と「後続音素」との間の歪みが音質に与える影響が大きい場合、「現音素」に対する境界周波数を決定するために「第2方式」が選択される。例えば、「現音素」が母音または有声子音を表し、「現音素」の直後の「後続音素」が母音または有声子音を表す場合に、「現音素」に対して「第2方式」が選択される。これによって境界周波数を高くし、「現音素」に対応する合成音声に占める素片接続型音声合成方式によるスペクトルを減少させ、「現音素」と「後続音素」との間の歪みに基づく異音発生を抑制する。

【0016】

例えば、「現音素」と「現音素」の直前の「先行音素」との間の歪みが音質に与える影響が小さく、かつ、「現音素」と「現音素」の直後の「後続音素」との間の歪みが音質に与える影響が小さい場合、「現音素」に対して「第3方式」を選択する。例えば、「現音素」が母音または有声子音を表し、「現音素」の直前の「先行音素」および直後の「後続音素」が無声子音を表す場合、および/または、現音素が無声子音を表す場合に、「現音素」に対して「第3方式」が選択される。このように歪みが音質に与える影響が小さい場合、境界周波数を低くし、「現音素」に対応する合成音声に占める素片接続型音声合成方式によるスペクトルを増加させる。これにより、歪みに基づく異音発生を抑えつつ、合成音声の自然感を向上させる。

【0017】

「第1方式」で、所定の周波数区間の一部の帯域である「第1判定帯域」での前述の「スペクトル特徴量の距離」を得、「第1判定帯域」での「スペクトル特徴量の距離」が「許容限界値(閾値)」未満であれば、「第1判定帯域」に応じた周波数を境界周波数とし

10

20

30

40

50

、「第1判定帯域」での「スペクトル特徴量の距離」が「許容限界値」未満でなければ、「第1判定帯域」よりも周波数の高い帯域を「第2判定帯域」とし、「第2判定帯域」を「第1判定帯域」とした同様な処理を再び実行してもよい。すなわち、低域側の「第1判定帯域」から順次「スペクトル特徴量の距離」を計算し、「スペクトル特徴量の距離」が「許容限界値」未満であれば、そのときの「第1判定帯域」に応じた周波数を境界周波数としてもよい。「第1判定帯域に応じた周波数」の例は、「第1判定帯域」の下限周波数、「第1判定帯域」の上限周波数、または「第1判定帯域」の中心周波数などである。「第2判定帯域」の例は、「第1判定帯域」の上限周波数または当該上限周波数に隣接する周波数を下限周波数とする帯域、あるいは「第1判定帯域」の上限周波数よりも高いその他の周波数を下限周波数とする帯域である。「第1判定帯域」の初期値の例は、所定の周波数区間の下限周波数を下限とする帯域である。人間の聴覚特性上、周波数が高いほど隣接する音素の接続部での歪みに基づく異音が合成音声の品質に与える影響が小さい。そのため、「第1判定帯域」での「スペクトル特徴量の距離」が「許容限界値」未満であって音素間の歪みが小さいのであれば、「第1判定帯域」よりも周波数が高い帯域でも音素間の歪みに基づく異音が合成音声の品質に与える影響も小さいことが多い。そのため、「スペクトル特徴量の距離」が「許容限界値」未満となった「第1判定帯域」に応じた周波数を境界周波数とすることで、歪みに基づく異音の影響を抑えつつ、合成音声の自然感を向上できる。また、「スペクトル特徴量の距離」が「許容限界値」未満となるまで、「第1判定帯域」の周波数を上げながら上述の処理を繰り返すことにより、歪みに基づく異音の影響を抑制可能なだけ低い周波数を境界周波数とできる。その結果、歪みに基づく異音の影響を抑えつつ、合成音声に占める素片接続型音声合成方式によるスペクトルをできるだけ増加させ、合成音声の自然感を向上できる。

10

20

30

40

50

**【0018】**

「許容限界値」は、「第1判定帯域」ごと（第1判定帯域に応じた周波数ごと）に定められてもよいし、すべての「第1判定帯域」に対して均一であってもよい。「許容限界値」が「第1判定帯域」ごとに定められる場合、「許容限界値」が「第1判定帯域」に応じた周波数に対して広義単調増加する関係にあってもよい。例えば「第1判定帯域」に応じた周波数が高いほど「許容限界値」が大きくてもよい。この場合、「許容限界値」が大きいほど、「第1判定帯域」での「スペクトル特徴量の距離」が「許容限界値」未満となる頻度が上がり、低い周波数が境界周波数として選択される頻度が高くなる。一方、周波数が高いほど隣接する音素の接続部での歪みに基づく異音が合成音声の品質に与える影響は小さくなる。そのため、「許容限界値」が「第1判定帯域」に応じた周波数に対して広義単調増加する関係にある場合、歪みに基づく異音の影響を抑えつつ、合成音声に占める素片接続型音声合成方式によるスペクトルをできるだけ増加させ、合成音声の自然感を向上させることができる。

**【0019】**

「第2方式」および「第3方式」が、「テキスト」に応じて素片接続される音素の基本周波数に対する、「テキスト」に前述の「音響モデル」を適用して得られる基本周波数の「変更度合い」に基づいて境界周波数を決定する方式であってもよい。「第2方式」および「第3方式」で選択される「境界周波数」が「変更度合い」の大きさに対して広義単調増加する関係にあってもよい。例えば、「第2方式」において、「変更度合い」が所定の範囲内である場合に「所定の周波数区間」の上限値を境界周波数とし、そうでない場合に「所定の周波数区間」の上限値を超える周波数（例えば、ナイキスト周波数）を境界周波数としてもよい。「第3方式」において、「変更度合い」が所定の範囲内である場合に「所定の周波数区間」の下限値未満の周波数（例えば、0 Hz）を境界周波数とし、そうでない場合に「所定の周波数区間」の下限値を境界周波数としてもよい。一般に「変更度合い」が大きいほど、素片接続型音声合成方式で得られる合成音声の品質は低下する。そのため、「変更度合い」に基づいて境界周波数を決定することで、合成音声の品質を向上させることができる。

**【0020】**

### [ 第 1 実施形態 ]

第 1 実施形態を説明する。

#### < 構成 >

図 1 に例示するように、本形態の音声合成装置 1 は、入力部 1 1、音声コーパス記憶部 2 1 2、音声データベース (DB) 記憶部 1 2 2、音響モデル記憶部 1 2 3、音声データベース (DB) 構築部 1 3 1、音響モデル生成部 1 3 2、素片接続型音声合成部 1 3 3、HMM 音声合成部 1 3 4、境界周波数決定部 1 3 5、スペクトル混合処理部 1 3 6、および波形生成処理部 1 3 7 を有する。図 2 A に例示するように、本形態の境界周波数決定部 1 3 5 は決定部 1 3 5 2 を含む。図 2 B に例示するように、本形態のスペクトル混合処理部 1 3 6 は、ハイパスフィルタ 1 3 6 1、ローパスフィルタ 1 3 6 2、および混合部 1 3 6 3 を有する。実施形態の音声合成装置は、例えば、CPU (central processing unit) 等のプロセッサ (ハードウェア・プロセッサ) および RAM (random-access memory) ・ROM (read-only memory) 等のメモリ等を備える汎用または専用のコンピュータが所定のプログラムを実行することで構成される装置である。このコンピュータは 1 個のプロセッサやメモリを備えていてもよいし、複数個のプロセッサやメモリを備えていてもよい。このプログラムはコンピュータにインストールされてもよいし、予め ROM 等に記録されていてもよい。また、CPU のようにプログラムが読み込まれることで機能構成を実現する電子回路 (circuitry) ではなく、プログラムを用いることなく処理機能を実現する電子回路を用いて一部またはすべての処理部が構成されてもよい。また、1 個の装置を構成する電子回路が複数の CPU を含んでいてもよい。

10

20

#### 【 0 0 2 1 】

##### < 前処理 >

前処理として音声コーパス記憶部 1 2 1 に音声コーパスが格納される。音声 DB 構築部 1 3 1 は、音声コーパス記憶部 1 2 1 に格納された音声コーパスを用い、素片接続型音声合成方式による音声合成に利用可能な音声 DB を生成する。例えば、音声 DB 構築部 1 3 1 は、特許 2 7 6 1 5 5 2 号公報、特許 4 4 3 0 9 6 0 号公報などに記載された公知の方法を用いて音声 DB を生成できる。生成された音声 DB は、音声 DB 記憶部 1 2 2 に格納される。音響モデル生成部 1 3 2 には、音声コーパス記憶部 1 2 1 に格納された音声コーパスを学習データとして用い、音声合成のための音響モデルを生成する。音響モデルの例は、HMM 音声合成用の音響モデルであり、例えば非特許文献 1 などに記載された公知の方法を用いて生成できる。本形態の音響モデルは、HMM スペクトル (周波数スペクトル) および基本周波数等をモデル化したものである。生成された音響モデルは音響モデル記憶部 1 2 3 に格納される。

30

#### 【 0 0 2 2 】

##### < 音声合成処理 >

次に、図 3 を用いて本形態の音声合成処理を説明する。入力部 1 1 には音声合成の対象となる文章を表すテキストが入力される。テキストは素片接続型音声合成部 1 3 3 および HMM 音声合成部 1 3 4 に入力される。HMM 音声合成部 1 3 4 は、音響モデル記憶部 1 2 3 に格納された音響モデルにテキストを適用 (例えば、テキストのテキスト解析結果を適用) して HMM スペクトルおよび基本周波数等を得て出力する。HMM スペクトルおよび基本周波数等はスペクトル混合処理部 1 3 6 に送られ、基本周波数はさらに素片接続型音声合成部 1 3 3 に送られる (ステップ S 1 3 4 : HMM 音声合成処理)。素片接続型音声合成部 1 3 3 は、テキストおよび基本周波数を入力とし、音声 DB 記憶部 1 2 2 に格納された音声 DB を用いて当該テキストに応じた素片接続 (素片接続型音声合成) を行い、素片スペクトル (当該テキストに応じた素片接続によって得られる第 1 合成音声のスペクトル) およびそれに対応する時間情報付の音素系列を得て出力する。「時間情報付の音素系列」は、素片接続された音素の系列であって各音素の時間情報が付与されたものである。音素の時間情報とは、例えば、所定の時刻 (例えば、音素系列の先頭時刻) を基準とした音素の時間軸上の位置 (例えば、音素の開始時刻または終了時刻) を表す情報である。素片スペクトルおよびそれに対応する時間情報付の音素系列は境界周波数決定部 1 3 5 に

40

50

送られ、素片スペクトルはさらにスペクトル混合処理部 1 3 6 に送られる（ステップ S 1 3 3：素片接続型音声合成処理）。

【 0 0 2 3 】

境界周波数決定部 1 3 5 は、素片スペクトルおよび時間情報付の音素系列を入力とし、前述のようにテキストに応じて素片接続された音素の素片境界前後のスペクトル特徴量の距離に基づいて境界周波数 B を決定して出力する（ステップ S 1 3 5：境界周波数決定処理）。

【 0 0 2 4 】

ステップ S 1 3 5 の詳細の例示

図 4 A および図 4 B を用い、ステップ S 1 3 5 の詳細を例示する。本形態では素片接続された音素ごとに境界周波数を決定する。本形態では、境界周波数を決定しようとする現音素（ $i$  番目の音素）と当該現音素の直前の先行音素（ $i - 1$  番目の音素）との音素境界  $T_i$  の前後の計算区間長  $T$  の時間区間における「所定の周波数区間」が  $N$  個（たとえば 10 個）の帯域（周波数帯域） $b_1, \dots, b_N$  に区分される。ただし、 $T$  は正値であり、例えば 20 msec である。 $N$  は 2 以上の正整数であり、例えば  $T = 10$  である。「所定の周波数区間」の上限値  $B_N$  および下限値  $B_0$  は予め定められている。上限値  $B_N$  の例は母音の第 2 フォルマントの上限周波数（例えば 3 kHz）であり、下限値  $B_0$  の例は母音の第 1 フォルマントの下限周波数（例えば 200 Hz）である。各帯域  $b_n$  の上限値を  $B_n$  と表記する。本形態では  $n$  が大きな帯域  $b_n$  ほど周波数が高く、帯域  $b_{n-1}$  の上限値  $B_{n-1}$  が帯域  $b_n$  の下限値と一致する。帯域  $b_1, \dots, b_N$  は「所定の周波数区間」をメル尺度（メル周波数）上で等間隔に区分したものであることが望ましいが、線形尺度（線形周波数）上で等間隔に区分したものの等、それ以外の基準で区分されたものであってもよい。

10

20

【 0 0 2 5 】

境界周波数決定部 1 3 5 の決定部 1 3 5 2（図 2 A）は  $n := 1$  と初期化する。ただし、「 $n := 1$ 」は  $n$  を 1 とする（ $n$  に 1 を代入する）ことを意味する（ステップ S 1 3 5 2 a）。決定部 1 3 5 2 は、先行音素の  $T_{i-1}$  から  $T_i$  までの時間区間における帯域  $b_n$ （所定の周波数区間の一部の帯域である第 1 判定帯域）でのスペクトル特徴量の平均値  $S_{n,i-1}$  と、現音素の  $T_i$  から  $T_i + T$  までの時間区間における帯域  $b_n$  でのスペクトル特徴量の平均値  $S_{n,i}$  とを計算する。例えば、帯域  $b_n$  での  $i - 1$  番目の音素（先行音素）の離散時刻  $t$  でのスペクトル特徴量を  $s_{n,i-1,t}$  とし、帯域  $b_n$  での  $i$  番目の音素（現音素）の離散時刻  $t$  でのスペクトル特徴量を  $s_{n,i,t}$  とすると、以下の関係が成り立つ。

30

【数 1】

$$S_{n,i-1} = \frac{1}{T} \sum_{t=T_{i-1}}^{T_i} s_{n,i-1,t}$$

$$S_{n,i} = \frac{1}{T} \sum_{t=T_i}^{T_i+T} s_{n,i,t}$$

40

なお、スペクトル特徴量の例は、パワースペクトル、メルケプストラム係数、ケプストラム係数などである（ステップ S 1 3 5 2 b）。

【 0 0 2 6 】

次に決定部 1 3 5 2 は、 $S_{n,i-1}$  と  $S_{n,i}$  との距離  $D_{n,i}$ （第 1 判定帯域  $b_n$  でのスペクトル特徴量の距離）を計算する。 $S_{n,i-1}$  および  $S_{n,i}$  がスカラーである場合、距離  $D_{n,i}$  は  $S_{n,i-1}$  と  $S_{n,i}$  との差分（絶対値）であり、 $S_{n,i-1}$  および  $S_{n,i}$  がベクトルである場合、距離  $D_{n,i}$  は  $S_{n,i-1}$  と  $S_{n,i}$  とのノルムである（ステップ S 1 3 5 2 c）。決定部 1 3 5 2 は、帯域  $b_n$  における許容限界値

50

$L_n$  と距離  $D_{n,i}$  とを比較し、 $L_n > D_{n,i}$  であるかを判定する。ただし、許容限界値  $L_n$  は聴取実験などによって予め定めておいた値である。許容限界値  $L_n$  は、帯域  $b_n$  ごとに定められてもよいし、すべての帯域  $b_1, \dots, b_N$  に対して均一であってもよい。許容限界値  $L_n$  が帯域  $b_n$  ごとに定められる場合、許容限界値  $L_n$  が帯域  $b_n$  に応じた周波数（例えば、 $B_n$  または  $B_{n-1}$ ）に対して広義単調増加する関係にあってもよい（ステップ S 1 3 5 2 d）。ここで、距離  $D_{n,i}$  が許容限界値  $L_n$  未満（ $L_n > D_{n,i}$ ）であれば、決定部 1 3 5 2 は、帯域  $b_n$  に応じた周波数を境界周波数  $B$  とする。例えば、 $B = B_n$  としてもよいし、 $B = B_{n-1}$  としてもよいし、 $B = (B_{n-1} + B_n) / 2$  としてもよい（ステップ S 1 3 5 2 g）。一方、距離  $D_{n,i}$  が許容限界値  $L_n$  未満でなければ、決定部 1 3 5 2 は  $n = N$  であるかを判定する（ステップ S 1 3 5 2 e）。ここで、 $n = N$  でなければ、決定部 1 3 5 2 は、 $n := n + 1$  として（ステップ S 1 3 5 2 f）、ステップ S 1 3 5 2 b 以降の処理を実行する。すなわち決定部 1 3 5 2 は、第 1 判定帯域（帯域  $b_n$ ）でのスペクトル特徴量の距離  $D_{n,i}$  が許容限界値  $L_n$  未満でなければ、第 1 判定帯域よりも周波数の高い帯域  $b_{n+1}$ （この例では 1 つ高域側の帯域）を第 2 判定帯域とし、第 2 判定帯域を第 1 判定帯域とした処理を行う。一方、 $n = N$  であればステップ S 1 3 5 2 g が実行され、帯域  $b_N$  に応じた周波数（例えば上限値  $B_N$ ）を境界周波数  $B$  とする（ステップ S 1 3 5 2 g）。以上のステップ S 1 3 5 2 a ~ S 1 3 5 2 g の処理は、音素接続される各音素を現音素としてそれぞれ実行される。このような処理により、定量的な基準に基づいて各音素に対応する境界周波数  $B$  を決定できる。

10

20

30

40

50

#### 【0027】

境界周波数決定部 1 3 5 から出力された境界周波数  $B$  はスペクトル混合処理部 1 3 6 に送られる。スペクトル混合処理部 1 3 6 は、音素ごとに、入力された素片スペクトル（テキストに応じた素片接続によって得られる第 1 合成音声のスペクトル）の境界周波数  $B$  に応じた高域側の成分と、HMM スペクトル（テキストに音声合成のための音響モデルを適用して得られる第 2 合成音声のスペクトル）の境界周波数  $B$  に応じた低域側の成分とを混合した混合スペクトルを得る。図 5 の例では、素片スペクトルおよび境界周波数  $B$  がハイパスフィルタ 1 3 6 1 に入力される。ハイパスフィルタ 1 3 6 1 は、ハイパスフィルタ処理によって素片スペクトルの境界周波数  $B$  以下の低域部をカットし、素片スペクトルの高域側の成分を得て出力する。また HMM スペクトルおよび境界周波数  $B$  がローパスフィルタ 1 3 6 2 に入力される。ローパスフィルタ 1 3 6 2 は、ローパスフィルタ処理によって HMM スペクトルの境界周波数  $B$  以上の高域部をカットすることにより HMM スペクトルの低域側の成分を得て出力する。混合部 1 3 6 3 は、素片スペクトルの高域側の成分および HMM スペクトルの低域側の成分を入力とし、これらを混合（合成）して混合スペクトルを得て出力する。これらの処理は全て音素毎に行う。混合スペクトルは基本周波数とともに波形生成処理部 1 3 7 に送られる（ステップ S 1 3 6 : スペクトル混合処理）。

#### 【0028】

波形生成処理部 1 3 7 は、入力された混合スペクトルおよび基本周波数を用い、混合スペクトルに対応する時間領域の波形（合成音声）を生成して出力する。この処理には、例えば、参考文献 1（H.Kawahara, "STRAIGHT, exploitation of the other aspect of Vocoder: Perceptually isomorphic decomposition of speech sounds", Acoustic Science and Technology, Vol.27, No.6, pp.349-353, 2006）に記載された方法を用いることができる（ステップ S 1 3 7 : 波形生成処理）。

#### 【0029】

##### [第2実施形態]

本形態は第 1 実施形態の変形例である。本形態では、テキストに応じて素片接続される音素の種別に応じ、(a) スペクトル特徴量の距離に基づいて所定の周波数区間の境界周波数を決定する第 1 方式、(b) 周波数区間の上限値以上の周波数を境界周波数とする第 2 方式、または (c) 周波数区間の下限値以下の周波数を境界周波数とする第 3 方式の何れかを選択する。以下では既に説明した事項との相違点を中心に説明し、説明済みの事項についてはそれまでに用いた参照番号を流用して説明を省略する。

## 【 0 0 3 0 】

< 構成 >

図 1 に例示するように、本形態の音声合成装置 2 は、入力部 1 1、音声コーパス記憶部 2 1 2、音声 D B 記憶部 1 2 2、音響モデル記憶部 1 2 3、音声 D B 構築部 1 3 1、音響モデル生成部 1 3 2、素片接続型音声合成部 1 3 3、H M M 音声合成部 1 3 4、境界周波数決定部 2 3 5、スペクトル混合処理部 1 3 6、および波形生成処理部 1 3 7 を有する。図 2 A に例示するように、本形態の境界周波数決定部 2 3 5 は、決定方法選択部 2 3 5 1 および決定部 2 3 5 2 を含む。

## 【 0 0 3 1 】

< 前処理 >

第 1 実施形態と同じである。

## 【 0 0 3 2 】

< 音声合成処理 >

第 1 実施形態との相違点は、境界周波数決定部 1 3 5 がステップ S 1 3 5 の境界周波数決定処理を行うことに代えて、境界周波数決定部 2 3 5 がステップ S 2 3 5 の境界周波数決定処理を行うことである。その他は第 1 実施形態で説明した通りである。以下では、ステップ S 2 3 5 の境界周波数決定処理のみを説明する。

## 【 0 0 3 3 】

ステップ S 2 3 5

境界周波数決定部 2 3 5 は、素片スペクトルおよび時間情報付の音素系列を入力とし、前述のようにテキストに応じて素片接続された音素の素片境界前後のスペクトル特徴量の距離に基づいて境界周波数 B を決定して出力する（ステップ S 2 3 5：境界周波数決定処理）。図 6 を用いて本形態のステップ S 2 3 5 の処理を説明する。

## 【 0 0 3 4 】

まず、境界周波数決定部 2 3 5 の決定方法選択部 2 3 5 1（図 2 A）は、時間情報付の音素系列によって表された素片接続される音素の種別（音素そのものの種別および隣接する音素の種別の組み合わせ）に応じ、音素ごとに、スペクトル特徴量の距離に基づいて、方式 a（第 1 方式）、方式 b（第 2 方式）、または方式 c（第 3 方式）の何れかを選択する（図 7 B 参照）。

## 【 0 0 3 5 】

方式 a は「所定の周波数区間（下限値  $B_0$  から上限値  $B_N$  までの区間）」内の境界周波数 B を決定する方式であり、第 1 実施形態で説明した方式である。前述のように、方式 a では、テキストに応じて素片接続される音素のそれぞれである「現音素」と「現音素」の直前の「先行音素」とのスペクトル特徴量の距離に基づいて「現音素」に対する境界周波数 B を決定する（例えば、図 4 A）。すなわち方式 a は「現音素」と「先行音素」との間の歪みが大きいときには素片接続型音声合成方式に適さないが、「現音素」と「先行音素」との間の歪みが小さいときには素片接続型音声合成方式を適用可能な「現音素」に向いている。決定方法選択部 2 3 5 1 は、このような「現音素」に対して方式 a を選択すればよい。

## 【 0 0 3 6 】

方式 b は「所定の周波数区間」の上限値  $B_N$  以上の周波数を境界周波数 B とする方式である。方式 b で選択される境界周波数 B の上限値は例えばナイキスト周波数である。この場合には、上限値  $B_N$  からナイキスト周波数までの間の境界周波数 B が選択される。方式 b で選択される境界周波数 B は方式 a で選択される境界周波数以上となる。すなわち方式 b は、基本的に素片接続型音声合成方式に適さない「現音素」、および隣接する音素とそのような関係を持った「現音素」に向いている。決定方法選択部 2 3 5 1 は、このような「現音素」に対して方式 b を選択すればよい。

## 【 0 0 3 7 】

方式 c は「所定の周波数区間」の下限値  $B_0$  以下の周波数を境界周波数 B とする方式である。方式 c で選択される境界周波数 B の下限値は例えば 0 H z である。この場合には、

10

20

30

40

50

0 Hz から下限値  $B_0$  までの間の境界周波数  $B$  が選択される。方式 c で選択される境界周波数  $B$  は方式 a で選択される境界周波数以下となる。すなわち方式 c は、基本的に素片接続型音声合成方式に適した「現音素」、および隣接する音素とそのような関係を持った「現音素」に向いている。決定方法選択部 2351 は、このような「現音素」に対して方式 c を選択すればよい。

#### 【0038】

方式の選択方法の例示 1

方式の選択方法を例示する。

(a) 決定方法選択部 2351 は、「現音素」と「現音素」の直前の「先行音素」との間の歪みが音質に与える影響が大きく、かつ、「現音素」と「現音素」の直後の「後続音素」との間の歪みが音質に与える影響が小さい場合に、「現音素」に対して方式 a を選択する。

10

(b) 決定方法選択部 2351 は、「現音素」と「現音素」の直後の「後続音素」との間の歪みが音質に与える影響が大きい場合に、「現音素」に対して方式 b を選択する。

(c) 決定方法選択部 2351 は、「現音素」と「現音素」の直前の「先行音素」との間の歪みが音質に与える影響が小さく、かつ、「現音素」と「現音素」の直後の「後続音素」との間の歪みが音質に与える影響が小さい場合に、「現音素」に対して方式 c を選択する。

#### 【0039】

方式の選択方法の例示 2

より具体的な方式の選択方法を例示する(図 7A)。

20

(a) 決定方法選択部 2351 は、「現音素」が母音または有声子音を表し、「現音素」の直前の「先行音素」が母音または有声子音を表し、「現音素」の直後の「後続音素」が無声子音を表す場合に、「現音素」に対して方式 a を選択する。

(b) 決定方法選択部 2351 は、「現音素」が母音または有声子音を表し、「現音素」の直後の「後続音素」が母音または有声子音を表す場合に、「現音素」に対して方式 b を選択する。

(c) 決定方法選択部 2351 は、「現音素」が母音または有声子音を表し、「現音素」の直前の「先行音素」および直後の「後続音素」が無声子音を表す場合、および / または、「現音素」が無声子音を表す場合に、「現音素」に対して方式 c を選択する。

30

#### 【0040】

このような境界周波数の決定方法の切り替えにより、音素の種別に基づいた最適な境界周波数の決定が可能となる。選択された方式を特定する情報、素片スペクトル、および時間情報付の音素系列は決定部 2352 に送られる(ステップ S2351)。

#### 【0041】

方式 a を特定する情報が決定部 2352 に送られた場合、決定部 2352 は第 1 実施形態のステップ S135 で説明した処理によって「現音素」に対応する境界周波数  $B$  を決定して出力する(ステップ S2352a)。方式 b を特定する情報が決定部 2352 に送られた場合、決定部 2352 は、上限値  $B_N$  以上の周波数(例えばナイキスト周波数)を「現音素」に対応する境界周波数  $B$  として決定して出力する(ステップ S2532b)。方式 c を特定する情報が決定部 2352 に送られた場合、決定部 2352 は、下限値  $B_0$  以下の周波数(例えば、0 Hz)を「現音素」に対応する境界周波数  $B$  として決定して出力する(ステップ S2532c)。以降の処理は第 1 実施形態と同じである。

40

#### 【0042】

[第 3 実施形態]

本形態は第 2 実施形態の変形例である。本形態では、方式 b (第 2 方式) および方式 c (第 3 方式) は、テキストに応じて素片接続される音素の基本周波数に対する、「第 1 合成音声」の基本周波数の変更度合いに基づいて境界周波数  $B$  を決定する。

#### 【0043】

<構成>

50

図 1 に例示するように、本形態の音声合成装置 3 は、入力部 1 1、音声コーパス記憶部 2 1 2、音声 DB 記憶部 1 2 2、音響モデル記憶部 1 2 3、音声 DB 構築部 1 3 1、音響モデル生成部 1 3 2、素片接続型音声合成部 1 3 3、HMM 音声合成部 1 3 4、境界周波数決定部 3 3 5、スペクトル混合処理部 1 3 6、および波形生成処理部 1 3 7 を有する。図 2 A に例示するように、本形態の境界周波数決定部 3 3 5 は、決定方法選択部 2 3 5 1 および決定部 3 3 5 2 を含む。

【 0 0 4 4 】

< 前処理 >

第 1 実施形態と同じである。

【 0 0 4 5 】

< 音声合成処理 >

第 2 実施形態との相違点は、境界周波数決定部 2 3 5 がステップ S 2 3 5 の境界周波数決定処理を行うことに代えて、境界周波数決定部 3 3 5 がステップ S 3 3 5 の境界周波数決定処理を行うことである。ステップ S 3 3 5 のステップ S 2 3 5 との相違点は、決定部 2 3 5 2 に代えて決定部 3 3 5 2 が以下の処理を行うことである。その他は第 1, 2 実施形態で説明した通りである。以下では決定部 3 3 5 2 の処理のみを説明する。

【 0 0 4 6 】

本形態では、決定方法選択部 2 3 5 1 で選択された方式を特定する情報、素片スペクトル、および時間情報付の音素系列が決定部 3 3 5 2 に送られる。さらに、決定部 3 3 5 2 には、ステップ S 1 3 4 で HMM 音声合成部 1 3 4 から出力された基本周波数も送られる。方式 a を特定する情報が決定部 3 3 5 2 に送られた場合、決定部 3 3 5 2 は第 1 実施形態のステップ S 1 3 5 で説明した処理によって「現音素」に対応する境界周波数 B を決定して出力する（ステップ S 2 3 5 2 a）。

【 0 0 4 7 】

方式 b を特定する情報が決定部 3 3 5 2 に送られた場合、決定部 3 3 5 2 は、「現音素」の音素区間における音素の基本周波数の平均値  $F_{0org}$  と、当該音素区間における HMM 音声合成部 1 3 4 での合成音の基本周波数の平均値  $F_{0syn}$  との比率  $F_{0org} / F_{0syn}$ （基本周波数の変更度合い）が予め定めた値（たとえば 0.5 以上 2.0 以下）の範囲を超えるかを判定する。基本周波数の平均値  $F_{0syn}$  は HMM 音声合成部 1 3 4 から出力された基本周波数に基づいて定められる（ステップ S 3 3 5 2 b a）。比率  $F_{0org} / F_{0syn}$  が予め定めた値の範囲内の場合、決定部 3 3 5 2 は境界周波数  $B = B_N$  として出力する（ステップ S 3 3 5 2 b b）。一方、比率  $F_{0org} / F_{0syn}$  が予め定めた値の範囲内でない場合、決定部 3 3 5 2 は、前述のステップ S 2 3 5 2 b によって境界周波数 B を決定して出力する（ステップ S 2 3 5 2 b）。

【 0 0 4 8 】

方式 c を特定する情報が決定部 3 3 5 2 に送られた場合も、決定部 3 3 5 2 は比率  $F_{0org} / F_{0syn}$  が予め定めた値の範囲を超えるかを判定する（ステップ S 3 3 5 2 c a）。比率  $F_{0org} / F_{0syn}$  が予め定めた値の範囲内の場合、決定部 3 3 5 2 は前述のステップ S 2 3 5 2 c によって境界周波数 B を決定して出力する（ステップ S 2 3 5 2 c）。一方、比率  $F_{0org} / F_{0syn}$  が予め定めた値の範囲内でない場合、決定部 3 3 5 2 は、境界周波数  $B = B_0$  として出力する（ステップ S 2 3 5 2 c b）。以降の処理は第 1 実施形態と同じである。

【 0 0 4 9 】

[ 第 3 実施形態の変形例 ]

第 3 実施形態では、方式 b が選択され、かつ、比率  $F_{0org} / F_{0syn}$  が予め定めた値の範囲内の場合に  $B = B_N$  とし、方式 b が選択され、かつ、比率  $F_{0org} / F_{0syn}$  が予め定めた値の範囲内でない場合にナイキスト周波数を B とした。しかしながら、方式 b が選択された場合に、その他の基準に則って、比率  $F_{0org} / F_{0syn}$  に応じて  $B_N$  以上ナイキスト周波数以下の周波数が B とされてもよい。例えば、 $B_N$  以上ナイキスト周波数以下の範囲で、比率  $F_{0org} / F_{0syn}$  が予め定められた値（例えば 1）

10

20

30

40

50

に近いほど、 $B_N$ に近い周波数が $B$ とされてもよい。

【0050】

第3実施形態では、方式cが選択され、かつ、比率 $F_{0org} / F_{0syn}$ が予め定められた値の範囲の場合に $B = 0$ とし、方式cが選択され、かつ、比率 $F_{0org} / F_{0syn}$ が予め定められた値の範囲でない場合に $B = B_0$ とした。しかしながら、方式cが選択された場合に、その他の基準に則って、比率 $F_{0org} / F_{0syn}$ に応じて0以上 $B_0$ 以下の周波数が $B$ とされてもよい。例えば、0以上 $B_0$ 以下の範囲で、比率 $F_{0org} / F_{0syn}$ が予め定められた値（例えば1）に近いほど、0に近い周波数が $B$ とされてもよい。

【0051】

第3実施形態では、「基本周波数の変更度合い」として比率 $F_{0org} / F_{0syn}$ を用いたが、これに代えてその他の指標を用いてもよい。例えば、比率 $F_{0syn} / F_{0org}$ 、差分 $|F_{0org} - F_{0syn}|$ 、二乗誤差 $\{(F_{0org})^2 - (F_{0syn})^2\}$ 等を「基本周波数の変更度合い」として用いてもよい。

【0052】

第3実施形態では、ステップS3352caで比率 $F_{0org} / F_{0syn}$ が予め定められた値の範囲を超えた場合に境界周波数 $B$ を下限値 $B_0$ としたが、この場合に境界周波数 $B$ をナイキスト周波数としてもよい。

【0053】

[特徴]

以上のように各実施形態では、隣接する音素間の連続性に基づく定量的な尺度、音素の種別、基本周波数の変更度合等に応じて適切に境界周波数を設定できる。これにより、HMM音声合成方式および素片接続型音声合成方式それぞれの長所を生かし、音素の接続部における異音の発生を抑制しつつ、音素のスペクトルの有する肉声感を導入した合成音声を生成できる。

【0054】

また各実施形態では、予め設定した許容限界値、音素の種別、基本周波数の変更度合いなどに基づいて定量的に境界周波数を決定する。そのため、話者や音声合成対象のテキストを変更するたびに手作業で境界周波数を定める作業が不要となり、自動的に境界周波数を決定できる。

【0055】

[その他の変形例]

上述の各種の処理は、記載に従って時系列に実行されるのみならず、処理を実行する装置の処理能力あるいは必要に応じて並列的あるいは個別に実行されてもよい。その他、本発明の趣旨を逸脱しない範囲で適宜変更が可能であることはいうまでもない。

【0056】

上述の構成をコンピュータによって実現する場合、各装置が有すべき機能の処理内容はプログラムによって記述される。このプログラムをコンピュータで実行することにより、上記処理機能がコンピュータ上で実現される。この処理内容を記述したプログラムは、コンピュータで読み取り可能な記録媒体に記録しておくことができる。コンピュータで読み取り可能な記録媒体の例は、非一時的な(non-transitory)記録媒体である。このような記録媒体の例は、磁気記録装置、光ディスク、光磁気記録媒体、半導体メモリ等である。

【0057】

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

【0058】

このようなプログラムを実行するコンピュータは、例えば、まず、可搬型記録媒体に記

10

20

30

40

50

録されたプログラムもしくはサーバコンピュータから転送されたプログラムを、一旦、自己の記憶装置に格納する。処理の実行時、このコンピュータは、自己の記録装置に格納されたプログラムを読み取り、読み取ったプログラムに従った処理を実行する。このプログラムの別の実行形態として、コンピュータが可搬型記録媒体から直接プログラムを読み取り、そのプログラムに従った処理を実行することとしてもよく、さらに、このコンピュータにサーバコンピュータからプログラムが転送されるたびに、逐次、受け取ったプログラムに従った処理を実行することとしてもよい。サーバコンピュータから、このコンピュータへのプログラムの転送は行わず、その実行指示と結果取得のみによって処理機能を実現する、いわゆる A S P (Application Service Provider) 型のサービスによって、上述の処理を実行する構成としてもよい。

【0059】

上記実施形態では、コンピュータ上で所定のプログラムを実行させて本装置の処理機能が実現されたが、これらの処理機能の少なくとも一部がハードウェアで実現されてもよい。

【符号の説明】

【0060】

1, 2, 3 音声合成装置

【図1】

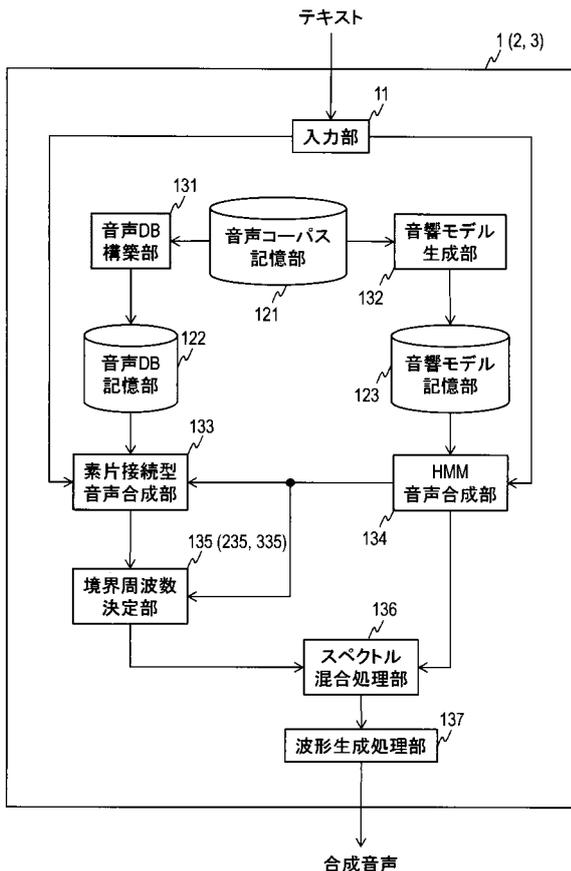


図1

【図2】

図2A

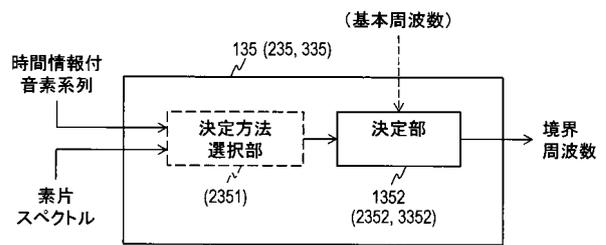
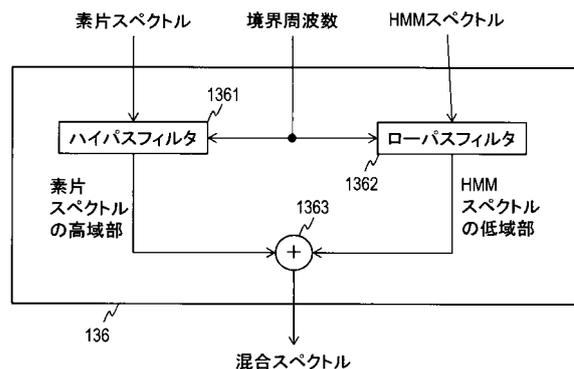
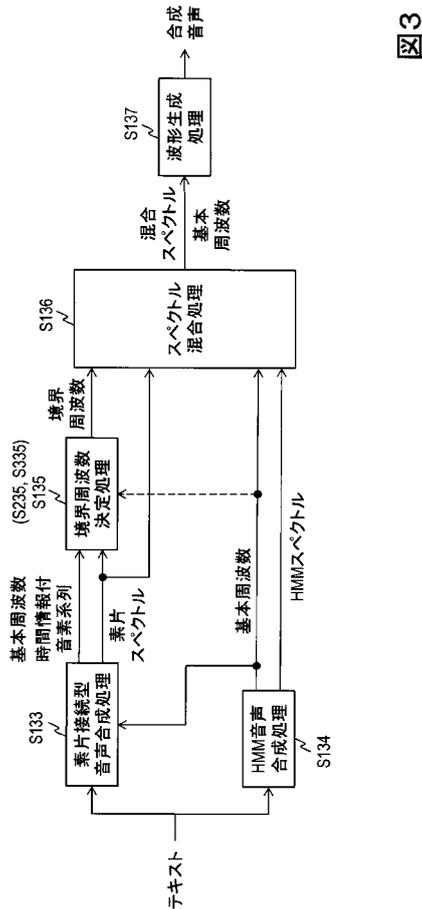


図2B



【 図 3 】



【 図 4 】

図4A

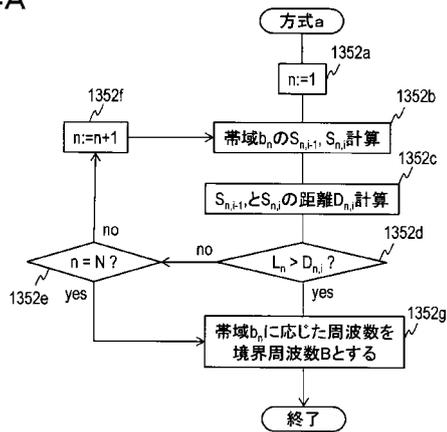
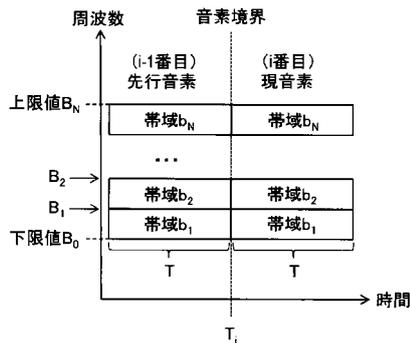
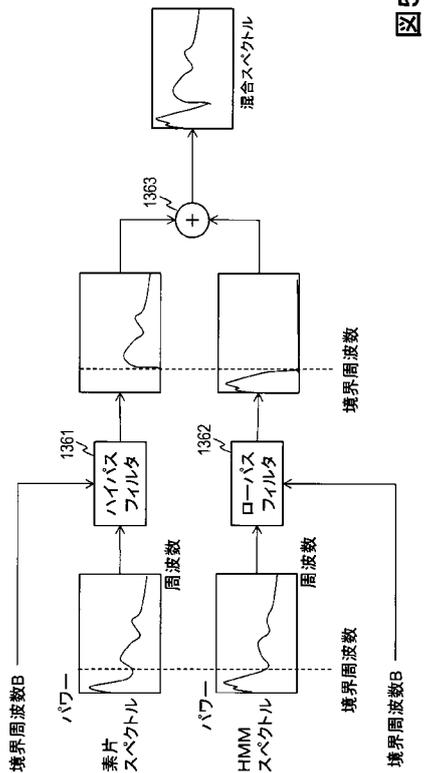


図4B



【 図 5 】



【 図 6 】

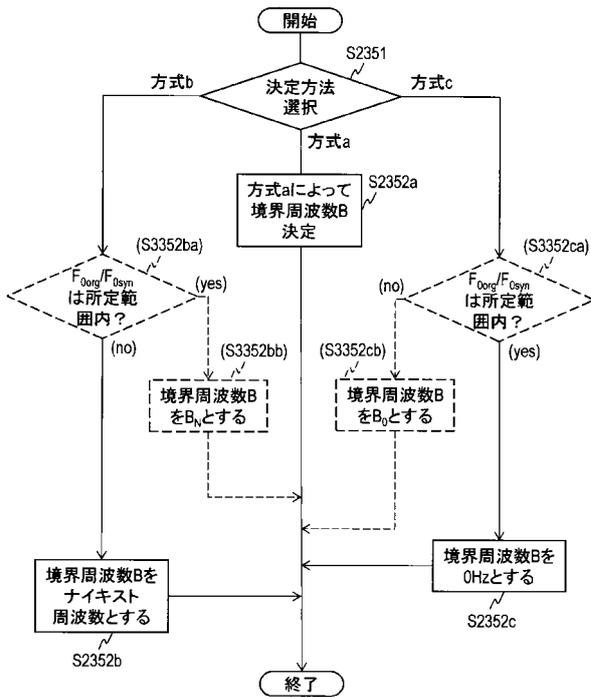


図6

## 【 図 7 】

図7A

先行音素	当該音素	後続音素	方式
—	母音または有 声子音	母音または有 声子音	方式b
母音または有 声子音	母音または有 声子音	無声子音	方式a
無声子音	母音または有 声子音	無声子音	方式c
—	無声子音	—	方式c

図7B

