

(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2012-163919

(P2012-163919A)

(43) 公開日 平成24年8月30日(2012.8.30)

(51) Int.Cl.

G10L 21/04 (2006.01)

F I

G10L 21/04 130A

テーマコード (参考)

審査請求 未請求 請求項の数 15 O L (全 33 頁)

(21) 出願番号 特願2011-26241 (P2011-26241)
(22) 出願日 平成23年2月9日(2011.2.9)

(71) 出願人 000002185
ソニー株式会社
東京都港区港南1丁目7番1号
(74) 代理人 100093241
弁理士 宮田 正昭
(74) 代理人 100101801
弁理士 山田 英治
(74) 代理人 100086531
弁理士 澤田 俊夫
(74) 代理人 100095496
弁理士 佐々木 榮二
(74) 代理人 110000763
特許業務法人大同特許事務所
(72) 発明者 光藤 祐基
東京都港区港南1丁目7番1号 ソニー株式会社内

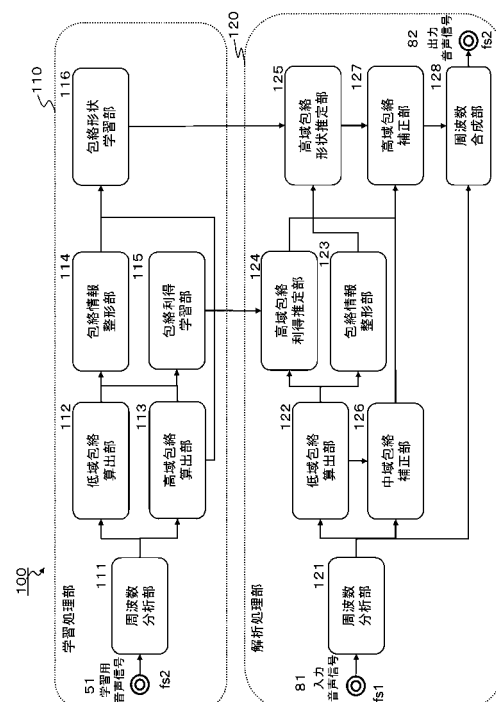
(54) 【発明の名称】 音声信号処理装置、および音声信号処理方法、並びにプログラム

(57) 【要約】

【課題】 音声信号に対する高精度な周波数帯域拡大処理を実行する装置、方法を提供する。

【解決手段】 入力音声信号の周波数分析結果に基づいて、低周波数帯域の包絡情報である低域包絡情報を算出する。さらに、学習用音声信号に基づく学習データ、例えば低域包絡情報から高周波数帯域の包絡情報である高域包絡情報を算出する学習データを適用して、入力音声信号対応の低域包絡情報から入力信号対応の高域包絡情報を推定して生成する。さらに、推定処理により生成した入力信号対応の高域包絡情報に対応する高周波数帯域信号を入力音声信号に合成して周波数帯域を拡大した出力音声信号を生成する。高域の包絡利得と包絡形状を、学習データを用いて推定することで高精度な帯域拡大が実現される。

【選択図】 図1



【特許請求の範囲】

【請求項 1】

入力音声信号の周波数分析を実行する周波数分析部と、

前記周波数分析部の周波数分析結果に基づいて、低周波数帯域の包絡情報である低域包絡情報を算出する低域包絡算出部と、

学習用音声信号に基づいて予め生成した学習データであり、低域包絡情報から高周波数帯域の包絡情報である高域包絡情報を算出するための学習データを適用して、前記入力音声信号対応の低域包絡情報から、入力信号対応の推定高域包絡情報を生成する高域包絡情報推定部と、

前記入力音声信号に、前記高域包絡情報推定部の生成した推定高域包絡情報に対応する高周波数帯域信号を合成して周波数帯域を拡大した出力音声信号を生成する周波数合成部を有する音声信号処理装置。

10

【請求項 2】

前記学習データは、

低域の包絡利得情報からの高域の包絡利得情報を推定するための包絡利得情報と、

低域の包絡形状情報からの高域の包絡形状情報を推定するための包絡形状情報を含み、

前記高域包絡情報推定部は、

前記学習データに含まれる包絡利得情報を適用して、前記入力音声信号対応の低域包絡利得情報から、入力信号対応の推定高域包絡利得情報を推定する高域包絡利得推定部と、

前記学習データに含まれる包絡形状情報を適用して、前記入力音声信号対応の低域包絡形状情報から、入力信号対応の推定高域包絡形状情報を推定する高域包絡形状推定部を有する請求項 1 に記載の音声信号処理装置。

20

【請求項 3】

前記高域包絡形状推定部は、

前記低域包絡算出部の算出した入力音声信号の低域包絡情報に対するフィルタリング処理により生成した整形低域包絡情報を入力して入力信号対応の推定高域包絡形状情報を推定する請求項 2 に記載の音声信号処理装置。

【請求項 4】

前記周波数分析部は、

前記入力音声信号の時間周波数分析を行い、時間周波数スペクトルを生成する請求項 1 に記載の音声信号処理装置。

30

【請求項 5】

前記低域包絡算出部は、

前記周波数分析部の生成する入力音声信号の時間周波数スペクトルを入力し、低域ケプストラムを生成する請求項 1 に記載の音声信号処理装置。

【請求項 6】

前記高域包絡情報推定部は、

前記学習データに含まれる包絡利得情報を適用して、前記入力音声信号対応の低域包絡利得情報から、入力信号対応の推定高域包絡利得情報を推定する高域包絡利得推定部を有し、

40

前記高域包絡利得推定部は、入力音声信号に基づいて生成された低域ケプストラム情報に対して、前記学習データに含まれる包絡利得情報を適用して、前記入力音声信号対応の低域包絡利得情報から、入力信号対応の推定高域包絡利得情報を推定する請求項 1 に記載の音声信号処理装置。

【請求項 7】

前記高域包絡情報推定部は、

前記学習データに含まれる包絡形状情報を適用して、前記入力音声信号対応の低域包絡形状情報から、入力信号対応の推定高域包絡形状情報を推定する高域包絡形状推定部を有し、

前記高域包絡形状推定部は、入力音声信号に基づいて生成された整形低域ケプストラム

50

情報に基づいて、前記学習データに含まれる包絡形状情報を利用した処理により、入力音声信号に対応する高域包絡形状情報の推定を行う請求項 1 に記載の音声信号処理装置。

【請求項 8】

前記高域包絡形状推定部は、GMM（ガウス混合モデル）を用いた推定処理により、入力音声信号に対応する高域包絡形状情報の推定を行う請求項 7 に記載の音声信号処理装置。

【請求項 9】

前記音声信号処理装置は、

前記入力音声信号に含まれない高周波数帯域の周波数を含む学習用音声信号に基づいて前記学習データを生成する学習処理部を有し、

前記高域包絡情報推定部は、

前記学習処理部の生成した学習データを適用して、前記入力音声信号対応の低域包絡情報から、入力信号対応の推定高域包絡情報を生成する請求項 1 に記載の音声信号処理装置。

【請求項 10】

第一の信号から第一の包絡情報を算出する機能と、

收音または伝送機能の少なくともいずれかを含む環境要因を排除する目的で第一の包絡情報の時間方向の DC 成分をフィルタによって除去する機能と、

フィルタ後の第一の包絡情報を線形変換した第二の包絡情報を第二の信号の包絡情報とし、第一の信号と合成する機能を備える音声信号処理装置。

【請求項 11】

低域の信号から低域の包絡情報を算出する機能と、

低域の包絡情報が、予め大量データの学習により分類されている複数のグループに属する確率を算出する機能と、

低域の包絡情報を、複数のグループにそれぞれ割り当てられている線形変換式によって線形変換を行い複数の高域包絡情報を生成する機能と、

時間軸において滑らかな高域包絡情報を生成する目的で複数の高域包絡情報を複数のグループに属する確率で混合することによって得られた高域の包絡情報を高域の信号の包絡情報とし、低域の信号と合成する機能を備える音声信号処理装置。

【請求項 12】

音声信号処理装置において入力音声信号に対する周波数帯域拡大処理を実行する音声信号処理方法であり、

周波数分析部が、入力音声信号の周波数分析を実行する周波数分析ステップと、

低域包絡算出部が、前記周波数分析ステップにおける周波数分析結果に基づいて、低周波数帯域の包絡情報である低域包絡情報を算出する低域包絡算出ステップと、

高域包絡情報推定部が、学習用音声信号に基づいて予め生成した学習データであり、低域包絡情報から高周波数帯域の包絡情報である高域包絡情報を算出するための学習データを適用して、前記入力音声信号対応の低域包絡情報から、入力信号対応の推定高域包絡情報を生成する高域包絡情報推定ステップと、

周波数合成部が、前記入力音声信号に、前記高域包絡情報推定部の生成した推定高域包絡情報に対応する高周波数帯域信号を合成して周波数帯域を拡大した出力音声信号を生成する周波数合成ステップを実行する音声信号処理方法。

【請求項 13】

音声信号処理装置において入力音声信号に対する処理を実行する音声信号処理方法であり、

第一の信号から第一の包絡情報を算出し、

收音または伝送機能の少なくともいずれかを含む環境要因を排除する目的で第一の包絡情報の時間方向の DC 成分をフィルタによって除去し、

フィルタ後の第一の包絡情報を線形変換した第二の包絡情報を第二の信号の包絡情報とし、第一の信号と合成する音声信号処理方法。

10

20

30

40

50

【請求項 14】

音声信号処理装置において入力音声信号に対する処理を実行する音声信号処理方法であり、

低域の信号から低域の包絡情報を算出し、

低域の包絡情報が、予め大量データの学習により分類されている複数のグループに属する確率を算出し、

低域の包絡情報を、複数のグループにそれぞれ割り当てられている線形変換式によって線形変換を行い複数の高域包絡情報を生成し、

時間軸において滑らかな高域包絡情報を生成する目的で複数の高域包絡情報を複数のグループに属する確率で混合することによって得られた高域の包絡情報を高域の信号の包絡情報とし、低域の信号と合成する音声信号処理方法。

10

【請求項 15】

音声信号処理装置において入力音声信号に対する周波数帯域拡大処理を実行させるプログラムであり、

周波数分析部に、入力音声信号の周波数分析を実行させる周波数分析ステップと、

低域包絡算出部に、前記周波数分析ステップにおける周波数分析結果に基づいて、低周波数帯域の包絡情報である低域包絡情報を算出させる低域包絡算出ステップと、

高域包絡情報推定部に、学習用音声信号に基づいて予め生成した学習データであり、低域包絡情報から高周波数帯域の包絡情報である高域包絡情報を算出するための学習データを適用して、前記入力音声信号対応の低域包絡情報から、入力信号対応の推定高域包絡情報を生成させる高域包絡情報推定ステップと、

20

周波数合成部に、前記入力音声信号に、前記高域包絡情報推定部の生成した推定高域包絡情報に対応する高周波数帯域信号を合成して周波数帯域を拡大した出力音声信号を生成させる周波数合成ステップを、

実行させるプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、音声信号処理装置、および音声信号処理方法、並びにプログラムに関する。さらに、詳細には入力信号に対する周波数帯域拡大処理を行う音声信号処理装置、および音声信号処理方法、並びにプログラムに関する。

30

【背景技術】

【0002】

データ通信やデータ記録処理に際しては、データ量の削減のために圧縮処理が行われることが多い。しかし、例えば音声信号を圧縮して伝送あるいは記録すると、オリジナルの音声データに含まれていた周波数帯域部分が失われる場合がある。

【0003】

従って、圧縮データを伸長して再生した場合、元のオリジナルの音声データとは異なる音声データとして再生されてしまうことがある。

このような圧縮データの伸長処理に際して、圧縮処理に際して失われた帯域部分を復元して伸長する構成についていくつか提案されている。

40

例えば特許文献1（特開2007-17908号公報）には、圧縮処理に際して失われた高域信号の生成処理を行う帯域拡大処理について開示している。

【0004】

しかし、従来の帯域拡大処理における問題点として、簡易な構成では高精度な拡大処理が困難となり、一方、高精度な拡大を実現するためには処理負荷や処理時間が増大し、装置としてのコストも大きくなるという問題点がある。

【先行技術文献】

【特許文献】

【0005】

50

【特許文献１】特開２００７－１７９０８号公報

【発明の概要】

【発明が解決しようとする課題】

【０００６】

本発明は、このような状況に鑑みてなされたものであり、簡易な構成で、より精度の高い帯域拡大処理を実現する音声信号処理装置、および音声信号処理方法、並びにプログラム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】

【０００７】

本発明の第１の側面は、

入力音声信号の周波数分析を実行する周波数分析部と、

前記周波数分析部の周波数分析結果に基づいて、低周波数帯域の包絡情報である低域包絡情報を算出する低域包絡算出部と、

学習用音声信号に基づいて予め生成した学習データであり、低域包絡情報から高周波数帯域の包絡情報である高域包絡情報を算出するための学習データを適用して、前記入力音声信号対応の低域包絡情報から、入力信号対応の推定高域包絡情報を生成する高域包絡情報推定部と、

前記入力音声信号に、前記高域包絡情報推定部の生成した推定高域包絡情報に対応する高周波数帯域信号を合成して周波数帯域を拡大した出力音声信号を生成する周波数合成部を有する音声信号処理装置にある。

【０００８】

さらに、本発明の音声信号処理装置の一実施態様において、前記学習データは、低域の包絡利得情報からの高域の包絡利得情報を推定するための包絡利得情報と、低域の包絡形状情報からの高域の包絡形状情報を推定するための包絡形状情報を含み、前記高域包絡情報推定部は、前記学習データに含まれる包絡利得情報を適用して、前記入力音声信号対応の低域包絡利得情報から、入力信号対応の推定高域包絡利得情報を推定する高域包絡利得推定部と、前記学習データに含まれる包絡形状情報を適用して、前記入力音声信号対応の低域包絡形状情報から、入力信号対応の推定高域包絡形状情報を推定する高域包絡形状推定部を有する。

【０００９】

さらに、本発明の音声信号処理装置の一実施態様において、前記高域包絡形状推定部は、前記低域包絡算出部の算出した入力音声信号の低域包絡情報に対するフィルタリング処理により生成した整形低域包絡情報を入力して入力信号対応の推定高域包絡形状情報を推定する。

【００１０】

さらに、本発明の音声信号処理装置の一実施態様において、前記周波数分析部は、前記入力音声信号の時間周波数分析を行い、時間周波数スペクトルを生成する。

【００１１】

さらに、本発明の音声信号処理装置の一実施態様において、前記低域包絡算出部は、前記周波数分析部の生成する入力音声信号の時間周波数スペクトルを入力し、低域ケプストラムを生成する。

【００１２】

さらに、本発明の音声信号処理装置の一実施態様において、前記高域包絡情報推定部は、前記学習データに含まれる包絡利得情報を適用して、前記入力音声信号対応の低域包絡利得情報から、入力信号対応の推定高域包絡利得情報を推定する高域包絡利得推定部を有し、前記高域包絡利得推定部は、入力音声信号に基づいて生成された低域ケプストラム情報に対して、前記学習データに含まれる包絡利得情報を適用して、前記入力音声信号対応の低域包絡利得情報から、入力信号対応の推定高域包絡利得情報を推定する。

【００１３】

さらに、本発明の音声信号処理装置の一実施態様において、前記高域包絡情報推定部は

10

20

30

40

50

、前記学習データに含まれる包絡形状情報を適用して、前記入力音声信号対応の低域包絡形状情報から、入力信号対応の推定高域包絡形状情報を推定する高域包絡形状推定部を有し、前記高域包絡形状推定部は、入力音声信号に基づいて生成された整形低域ケプストラム情報に基づいて、前記学習データに含まれる包絡形状情報を利用した処理により、入力音声信号に対応する高域包絡形状情報の推定を行う。

【0014】

さらに、本発明の音声信号処理装置の一実施態様において、前記高域包絡形状推定部は、GMM（ガウス混合モデル）を用いた推定処理により、入力音声信号に対応する高域包絡形状情報の推定を行う。

【0015】

さらに、本発明の音声信号処理装置の一実施態様において、前記音声信号処理装置は、前記入力音声信号に含まれない高周波数帯域の周波数を含む学習用音声信号に基づいて前記学習データを生成する学習処理部を有し、前記高域包絡情報推定部は、前記学習処理部の生成した学習データを適用して、前記入力音声信号対応の低域包絡情報から、入力信号対応の推定高域包絡情報を生成する。

【0016】

さらに、本発明の第2の側面は、

第一の信号から第一の包絡情報を算出する機能と、

收音または伝送機能の少なくともいずれかを含む環境要因を排除する目的で第一の包絡情報の時間方向のDC成分をフィルタによって除去する機能と、

フィルタ後の第一の包絡情報を線形変換した第二の包絡情報を第二の信号の包絡情報とし、第一の信号と合成する機能を備える音声信号処理装置にある。

【0017】

さらに、本発明の第3の側面は、

低域の信号から低域の包絡情報を算出する機能と、

低域の包絡情報が、予め大量データの学習により分類されている複数のグループに属する確率を算出する機能と、

低域の包絡情報を、複数のグループにそれぞれ割り当てられている線形変換式によって線形変換を行い複数の高域包絡情報を生成する機能と、

時間軸において滑らかな高域包絡情報を生成する目的で複数の高域包絡情報を複数のグループに属する確率で混合することによって得られた高域の包絡情報を高域の信号の包絡情報とし、低域の信号と合成する機能を備える音声信号処理装置にある。

【0018】

さらに、本発明の第4の側面は、

音声信号処理装置において入力音声信号に対する周波数帯域拡大処理を実行する音声信号処理方法であり、

周波数分析部が、入力音声信号の周波数分析を実行する周波数分析ステップと、

低域包絡算出部が、前記周波数分析ステップにおける周波数分析結果に基づいて、低周波数帯域の包絡情報である低域包絡情報を算出する低域包絡算出ステップと、

高域包絡情報推定部が、学習用音声信号に基づいて予め生成した学習データであり、低域包絡情報から高周波数帯域の包絡情報である高域包絡情報を算出するための学習データを適用して、前記入力音声信号対応の低域包絡情報から、入力信号対応の推定高域包絡情報を生成する高域包絡情報推定ステップと、

周波数合成部が、前記入力音声信号に、前記高域包絡情報推定部の生成した推定高域包絡情報に対応する高周波数帯域信号を合成して周波数帯域を拡大した出力音声信号を生成する周波数合成ステップを実行する音声信号処理方法にある。

【0019】

さらに、本発明の第5の側面は、

音声信号処理装置において入力音声信号に対する処理を実行する音声信号処理方法であり、

10

20

30

40

50

第一の信号から第一の包絡情報を算出し、

收音または伝送機能の少なくともいずれかを含む環境要因を排除する目的で第一の包絡情報の時間方向のDC成分をフィルタによって除去し、

フィルタ後の第一の包絡情報を線形変換した第二の包絡情報を第二の信号の包絡情報とし、第一の信号と合成する音声信号処理方法にある。

【0020】

さらに、本発明の第6の側面は、

音声信号処理装置において入力音声信号に対する処理を実行する音声信号処理方法であり、

低域の信号から低域の包絡情報を算出し、

低域の包絡情報が、予め大量データの学習により分類されている複数のグループに属する確率を算出し、

低域の包絡情報を、複数のグループにそれぞれ割り当てられている線形変換式によって線形変換を行い複数の高域包絡情報を生成し、

時間軸において滑らかな高域包絡情報を生成する目的で複数の高域包絡情報を複数のグループに属する確率で混合することによって得られた高域の包絡情報を高域の信号の包絡情報とし、低域の信号と合成する音声信号処理方法にある。

【0021】

さらに、本発明の第7の側面は、

音声信号処理装置において入力音声信号に対する周波数帯域拡大処理を実行させるプログラムであり、

周波数分析部に、入力音声信号の周波数分析を実行させる周波数分析ステップと、

低域包絡算出部に、前記周波数分析ステップにおける周波数分析結果に基づいて、低周波数帯域の包絡情報である低域包絡情報を算出させる低域包絡算出ステップと、

高域包絡情報推定部に、学習用音声信号に基づいて予め生成した学習データであり、低域包絡情報から高周波数帯域の包絡情報である高域包絡情報を算出するための学習データを適用して、前記入力音声信号対応の低域包絡情報から、入力信号対応の推定高域包絡情報を生成させる高域包絡情報推定ステップと、

周波数合成部に、前記入力音声信号に、前記高域包絡情報推定部の生成した推定高域包絡情報に対応する高周波数帯域信号を合成して周波数帯域を拡大した出力音声信号を生成させる周波数合成ステップを、

実行させるプログラムにある。

【0022】

なお、本発明のプログラムは、例えば、様々なプログラム・コードを実行可能な画像処理装置やコンピュータ・システムに対して、コンピュータ可読な形式で提供する記憶媒体、通信媒体によって提供可能なプログラムである。このようなプログラムをコンピュータ可読な形式で提供することにより、情報処理装置やコンピュータ・システム上でプログラムに応じた処理が実現される。

【0023】

本発明のさらに他の目的、特徴や利点は、後述する本発明の実施例や添付する図面に基づくより詳細な説明によって明らかになるであろう。なお、本明細書においてシステムとは、複数の装置の論理的集合構成であり、各構成の装置が同一筐体内にあるものには限らない。

【発明の効果】

【0024】

本発明の一実施例の構成によれば、音声信号に対する高精度な周波数帯域拡大処理を実行する装置、方法が実現される。

例えば、本発明の一実施例の構成においては、入力音声信号の周波数分析結果に基づいて、低周波数帯域の包絡情報である低域包絡情報を算出する。さらに、学習用音声信号に基づく学習データ、例えば低域包絡情報から高周波数帯域の包絡情報である高域包絡情報

10

20

30

40

50

を算出する学習データを適用して、入力音声信号対応の低域包絡情報から入力信号対応の高域包絡情報を推定して生成する。さらに、推定処理により生成した入力信号対応の高域包絡情報に対応する高周波数帯域信号を入力音声信号に合成して周波数帯域を拡大した出力音声信号を生成する。高域の包絡利得と包絡形状を、学習データを用いて推定することで高精度な帯域拡大が実現される。

【図面の簡単な説明】

【0025】

【図1】音声信号処理装置の構成例について説明する図である。

【図2】周波数分析処理と包絡情報算出処理について説明する図である。

【図3】音源によって包絡形状（正確には次数毎のケプストラム）の時間変動が異なる様子を示した図である。 10

【図4】音声信号の包絡形状にDC成分が含まれる場合と、含まれない場合の包絡形状の時間変動を示す図である。

【図5】包絡形状DC成分の時系列データを示す図である。

【図6】包絡形状のDC成分の周波数領域の様子を示した図である。

【図7】KmeansとGMMによるモデル化データを参照した包絡形状学習部における包絡形状の推定処理について説明する図である。

【図8】KmeansとGMMによるモデル化データを参照した高域包絡形状推定部の実行する高域包絡形状情報の推定処理について説明する図である。

【図9】（a）Kmeansと（b）GMMをそれぞれ使用した場合、写像元のデータがクラスタ境界を越えて変化すると、写像先のデータがどのように変化するかを説明する図である。 20

【発明を実施するための形態】

【0026】

以下、図面を参照しながら本発明の音声信号処理装置、および音声信号処理方法、並びにプログラムの詳細について説明する。説明は以下の項目に従って行う。

1. 本発明の音声信号処理装置の全体構成について

2. 信号処理装置の各構成部の処理について

2. 1. 周波数分析部について

2. 2. 低域包絡算出部について 30

2. 3. 高域包絡算出部について

2. 4. 包絡情報整形部について

2. 5. 包絡利得学習部と包絡形状学習部について

2. 6. 高域包絡形状推定部について

2. 7. 高域包絡利得推定部について

2. 8. 中域包絡補正部について

2. 9. 高域包絡補正部について

2. 10. 周波数合成部について

【0027】

[1. 本発明の信号処理装置の全体構成について] 40

まず、図1を参照して、この発明の実施の形態に係る信号処理装置の全体構成について説明する。

図1は、この発明の実施の形態に係る音声信号処理装置100の一例を示す図である。図1に示す音声信号処理装置100は、上段の学習処理部110と、下段の解析処理部120を有する。

【0028】

解析処理部120に入力する入力音声信号81の周波数帯域拡大処理を実行して、出力音声信号82として出力する。この解析処理部120において実行する周波数帯域拡大処理に際して、学習処理部110が学習用音声信号51に基づいて生成したデータを利用する。 50

学習処理部 110 は、学習用音声信号 51 を入力して、学習用音声信号 51 の解析により、例えば周波数包絡等の学習データを生成する。解析処理部 120 は、学習処理部 110 の生成した学習結果を利用して入力音声信号 81 の周波数帯域拡大処理を行う。

【0029】

図 1 に示すように、学習処理部 110 は、周波数分析部 111、低域包絡算出部 112、高域包絡算出部 113、包絡情報整形部 114、包絡利得学習部 115、包絡形状学習部 116 を有する。

また解析処理部 120 は、周波数分析部 121、低域包絡算出部 122、包絡情報整形部 123、高域包絡利得推定部 124、高域包絡形状推定部 125、中域包絡補正部 126、高域包絡補正部 127、周波数合成部 128 を有する。

10

【0030】

図 1 に示す学習処理部 110 が学習対象として入力する学習用音声信号 51 のサンプリング周波数 (f_{s2}) と、解析処理部 120 の出力信号、すなわち周波数帯域拡大処理を行った後の出力音声信号 82 のサンプリング周波数 (f_{s2}) は同じである。

これらの 2 つの信号のサンプリング周波数 (f_{s2}) は、解析処理部 120 の入力信号、すなわち周波数帯域拡大処理の対象となる入力音声信号 81 のサンプリング周波数 (f_{s1}) に対して 2 倍の値となっている。

なお、 f_{s1} 、 f_{s2} はそれぞれサンプリング周波数を示し

$$(f_{s2}) = 2 \times (f_{s1})$$

上記の対応関係である。

20

【0031】

すなわち、解析処理部 120 の入力する入力音声信号 81 のサンプリング周波数 (f_{s1}) は周波数帯域の圧縮された信号であり、解析処理部 120 は、この入力信号の周波数帯域の拡大処理を実行して 2 倍のサンプリング周波数 (f_{s2}) の出力音声信号 82 を生成して出力する。

解析処理部 120 はこの帯域拡大処理に際して、出力音声信号 82 のサンプリング周波数 (f_{s2}) と同じサンプリング周波数 (f_{s2}) に対する学習データを学習処理部 110 から取得してこの学習データを利用して高精度な周波数帯域拡大処理を実行する。

以下、各構成部の処理の詳細について説明する。

【0032】

30

[2. 信号処理装置の各構成部の処理について]

(2.1. 周波数分析部について)

図 1 に示すように、周波数分析部は、学習処理部 110 と解析処理部 120 の双方に設定されている。

図 1 に示す学習処理部 110 の周波数分析部 111 は、サンプリング周波数 (f_{s2}) の学習用音声信号 51 を入力し、この学習用音声信号 51 に対して周波数分析を行う。

また、解析処理部 120 の周波数分析部 121 は、周波数帯域拡大処理の対象となる入力音声信号 81 に対して時間周波数分析を行う。

【0033】

これらの周波数分析部 111 と、周波数分析部 121 において実行する時間周波数分析処理について、図 2 を参照して説明する。

40

周波数分析部 111 と、周波数分析部 121 は、入力音声信号に対する時間周波数分析を実行する。

【0034】

マイク等を介して入力する入力信号を x とする。図 2 の最上段に入力信号 x の例を示す。横軸が時間 (またはサンプル番号)、縦軸が振幅である。

学習処理部 110 の周波数分析部 111 に対する入力信号 x は、サンプリング周波数 (f_{s2}) の学習用音声信号 51 である。

また、解析処理部 120 の周波数分析部 121 に対する入力信号 x は、周波数帯域拡大処理の処理対象信号であるサンプリング周波数 (f_{s1}) の入力音声信号 81 である。

50

【0035】

まず、周波数分析部111と、周波数分析部121は、入力信号xから固定サイズのフレーム分割を行い、入力フレーム信号x(n, l)を得る。

図2のステップS101の処理である。

図2に示す例では、フレーム分割のサイズをNとし、各フレームのシフト量(sf)をフレームの大きさNの50%とし、各フレームを重複させた設定としている。

【0036】

さらに、入力フレーム信号x(n, l)に対して、所定の窓関数:wを乗算して窓関数適用信号wx(n, l)を得る。窓関数は例えばハニング窓の平方根を計算したものが適用可能である。

窓関数適用信号wx(n, l)は、以下の(式1)によって示される。

【0037】

【数1】

$$wx(n, l) = w_{ana}(n) * x(n, l)$$

{

x:入力信号

n:時間インデックス

l:フレーム番号

w_{ana}:窓関数

wx:窓関数適用信号

}

$$w_{ana}(n) = \left(0.5 - 0.5 * \cos\left(2\pi \frac{n}{N}\right) \right)^{0.5} \quad (N: \text{フレームの大きさ})$$

・・・・・・(式1)

【0038】

上記式(式1)において、

x:入力信号、

n:時間インデックス、n=0, ..., N-1、l=0, ..., L-1
(Nはフレームの大きさ)

l:フレーム番号、l=0, ..., L-1
(Lは全フレーム数)

w_{ana}:窓関数、

wx:窓関数適用信号、
である。

【0039】

なお、窓関数:w_{ana}は、上記の例ではハニング窓の平方根を計算した窓関数を適用しているが、この他、サイン窓などの窓関数も使用可能である。

フレームの大きさNは、例えば、0.02sec相当のサンプル数(N=サンプリング周波数fs*0.02)である。ただし、それ以外の大きさでも構わない。

また、図2に示す例では、フレームのシフト量(sf)はフレームの大きさ(N)の50%とし、各フレームを重複した設定としているが、それ以外のシフト量でも構わない。

【0040】

上記(式1)に従って得られた窓関数適用信号wx(n, l)に対して、以下に示す(式2)に従って時間周波数分析を行い、時間周波数スペクトルX_{ana}(k, l)を得る。

。

【 0 0 4 1 】

【 数 2 】

$$X_{ana}(k, l) = \sum_{n=0}^{M-1} wx(n, l) * \exp\left(-j2\pi \frac{k * n}{M}\right)$$

wx : 窓関数適用信号
 j : 純虚数
 M : DFTのポイント数
 k : 周波数インデックス
 X_{ana} : 周波数スペクトル

10

$$wx(n, l) = \begin{cases} wx(n, l) & n = 0, \dots, N-1 \\ 0 & n = N, \dots, M-1 \end{cases}$$

・・・ (式 2)

【 0 0 4 2 】

上記式 (式 2) において、

 wx : 窓関数適用信号、 j : 純虚数、 M : DFT (離散フーリエ変換) のポイント数、 k : 周波数インデックス、 X_{ana} : 時間周波数スペクトル、

である。

20

【 0 0 4 3 】

窓関数適用信号 $wx(n, l)$ に対する時間周波数分析処理としては、例えば、DFT (離散フーリエ変換) による周波数分析が適用可能である。その他、DCT (離散コサイン変換) や MDCT (修正離散コサイン変換) などの他の周波数分析を用いてもよい。また、必要であれば、DFT (離散フーリエ変換) のポイント数 M に合わせて適切に零詰めを行ってもよい。DFT のポイント数 M は、 N 以上の 2 のべき乗の値としているが、それ以外のポイント数でも構わない。

30

【 0 0 4 4 】

(2. 2. 低域包絡算出部について)

低域包絡算出部も、上記の周波数分析部と同様、図 1 に示すように学習処理部 110 と解析処理部 120 の双方に設定されている。

【 0 0 4 5 】

学習処理部 110 の低域包絡算出部 112 は、サンプリング周波数 (f_{s2}) の学習用音声信号 51 に対する周波数分析部 111 の分析結果として得られる時間周波数スペクトルから選択される低周波数帯域 (例えば $f_{s1}/2$ 未満) の周波数に相当するスペクトルに対する処理によって低域の包絡情報を算出する。

40

【 0 0 4 6 】

一方、解析処理部 120 の低域包絡算出部 122 は、サンプリング周波数 (f_{s1}) の入力音声信号 81 に対する周波数分析部 121 の分析結果として得られる時間周波数スペクトルから選択される低周波数帯域 (例えば $f_{s1}/2$ 未満) の周波数に相当するスペクトルに対する処理によって低域の包絡情報を算出する。

【 0 0 4 7 】

これら 2 つの低域包絡算出部 112 と、低域包絡算出部 122 は、処理対象が異なるのみで、実行する処理は同じである。すなわち、周波数分析部の分析結果として得られる時間周波数スペクトルから選択される低周波数帯域 (例えば $f_{s1}/2$ 未満) の周波数に相

50

当するスペクトルに対する処理によって低域の包絡情報を算出する。

以下、この処理について説明する。

【0048】

低域包絡算出部112, 122では、周波数分析部111, 121より供給された0以上fs1/2未満の周波数に相当する時間周波数スペクトルXana(k, l)からスペクトルの微細構造を取り除き、包絡情報を算出する。例えば以下の(式3)に従って、低域包絡情報に相当する低域ケプストラムClowを算出する。

【0049】

【数3】

$$C_{low}(i, l) = \frac{1}{M} \sum_{k=0}^{M-1} \log(|X_{ana}(k, l)|) * \exp\left(j2\pi \frac{i * k}{M}\right) \quad \left(\begin{array}{l} i: \text{ケプストラムインデックス} \\ C_{low}: \text{低域ケプストラム} \end{array} \right)$$

10

・・・・・・(式3)

【0050】

上記式(式3)において、

i: ケプストラムインデックス、

Clow: 低域ケプストラム、

である。

20

【0051】

なお、この低域包絡算出部112, 122の処理は、図2に示すステップS102、S103の処理に対応する。

図2に示すステップS102は、上記(式3)に従った各フレーム対応の低域包絡情報の算出処理である。

ステップS103は、上記(式3)に従って算出された各フレーム対応の低域包絡情報を、

周波数(周波数ビン)を行、

時間(フレーム)を列、

としたN行L列の行列として、行列の各要素を示したものである。

30

【0052】

上記(式3)に示すように、低域包絡算出部112, 122は、L F C C(線形周波数ケプストラム係数、以下ケプストラム)を算出し、低次項の係数のみを利用することにより低域包絡情報を取得する。

【0053】

なお、低域包絡算出部112, 122における低域包絡情報の算出処理は、上記のようなL F C C(線形周波数ケプストラム係数、以下ケプストラム)の適用処理に限らず、例えば、L P C C(線形予測ケプストラム係数)、M F C C(メル周波数ケプストラム係数)、P L P C C(知覚線形予測ケプストラム係数)などのその他のケプストラムや、その他の周波数包絡情報を用いた構成としてもよい。

40

【0054】

図1に示す上段の学習処理部110の低域包絡算出部112は、学習用音声信号51に対して、上記の(式3)に従って算出した低域ケプストラムClow(i, l)を、包絡利得学習部115と包絡情報整形部114へ供給する。

【0055】

また、図1の下段の解析処理部120の低域包絡算出部122は、入力音声信号81に対して、上記の(式3)に従って算出した低域ケプストラムClow(i, l)を、高域包絡利得推定部124と包絡情報整形部123へ供給する。

【0056】

50

(2. 3. 高域包絡算出部について)

次に、高域包絡算出部の処理について説明する。

高域包絡算出部は、図1に示すように学習処理部110に設けられている。

学習処理部110の高域包絡算出部113は、サンプリング周波数(f_s2)の学習用音声信号51に対する周波数分析部111の分析結果として得られる時間周波数スペクトルから選択される高周波数帯域(例えば $f_s1/2$ 以上 $f_s2/2$ 未満)の周波数に相当するスペクトルに対する処理によって高域の包絡情報を算出する。

【0057】

高域包絡算出部113は、周波数分析部111より供給された $f_s1/2$ 以上 $f_s2/2$ 未満の周波数に相当する時間周波数スペクトル $X_{ana}(k, l)$ からスペクトルの微細構造を取り除き、包絡情報を算出する。例えば以下の(式4)に従って、高域包絡情報に相当する高域ケプストラム C_{high} を算出する。

【0058】

【数4】

$$C_{high}(i, l) = \frac{1}{M} \sum_{k=0}^{M-1} \log(|X_{ana}(k, l)|) * \exp\left(j2\pi \frac{i * k}{M}\right) \quad \left(\begin{array}{l} i: \text{ケプストラムインデックス} \\ C_{high}: \text{高域ケプストラム} \end{array} \right)$$

..... (式4)

【0059】

上記式(式4)において、

i : ケプストラムインデックス、

C_{high} : 低域ケプストラム、
である。

【0060】

上記のように、本実施例においてはL F C C (線形周波数ケプストラム係数、以下ケプストラム)を算出し、低次項の係数のみを利用することにより包絡情報を取得している。ただし、高域包絡算出部113における高域包絡情報の算出に際しては、L F C C (線形周波数ケプストラム係数、以下ケプストラム)に限らず、例えば、L P C C (線形予測ケプストラム係数)、M F C C (メル周波数ケプストラム係数)、P L P C C (知覚線形予測ケプストラム係数)などのその他のケプストラムや、その他の周波数包絡情報を用いる構成としてもよい。

【0061】

図1に示す上段の学習処理部110の低域包絡算出部112は、学習用音声信号51に対して、上記(式4)に従って算出した高域ケプストラム $C_{high}(i, l)$ を、包絡情報整形部114と、包絡利得学習部115、および包絡形状学習部116へ供給する。

【0062】

(2. 4. 包絡情報整形部について)

包絡情報整形部は、図1に示すように学習処理部110と解析処理部120の双方に設定されている。

【0063】

学習処理部110の包絡情報整形部114は、サンプリング周波数(f_s2)の学習用音声信号51に基づいて、

低域包絡算出部112が生成した低域包絡情報、

を入力し、フィルタリング処理により包絡情報の整形を実行し、整形包絡情報を生成して包絡形状学習部116に提供する。

【0064】

一方、解析処理部120の包絡情報整形部123は、サンプリング周波数(f_s1)の

入力音声信号 8 1 に基づいて、

低域包絡算出部 1 2 2 が生成した低域包絡情報、

を入力し、この包絡情報のフィルタリング処理により包絡情報の整形を実行し、整形包絡情報を生成して高域包絡形状推定部 1 2 5 に提供する。

【0065】

より、具体的には、学習処理部 1 1 0 の包絡情報整形部 1 1 4 は、サンプリング周波数 (f_{s2}) の学習用音声信号 5 1 に基づいて、

低域包絡算出部 1 1 2 が生成した低域包絡情報、

すなわち、上記 (式 3) に従って算出した低域ケプストラム $C_{low}(i, l)$ を入力し、包絡情報: $C_{low}(i, l)$ 、に対して、予め定めた次数 R までを残し、以下を削除するフィルタリング処理を行う包絡情報の整形を実行し、整形包絡情報: $C'_{low}(i, l)$ を生成して包絡情報学習部 1 1 6 に提供する。

【0066】

一方、解析処理部 1 2 0 の包絡情報整形部 1 2 3 は、サンプリング周波数 (f_{s1}) の入力音声信号 8 1 に基づいて、

低域包絡算出部 1 2 2 が生成した低域包絡情報、すなわち、上記 (式 3) に従って算出した低域ケプストラム $C_{low}(i, l)$

を入力し、この包絡情報: $C_{low}(i, l)$ を、次数毎にフレーム方向にフィルタ処理を行い、変調周波数の DC 成分と 25 Hz 以上の高周波成分を除去する整形を実行し、整形包絡情報 ($C'_{low}(i, l)$) を生成して高域包絡形状推定部 1 2 5 に提供する。

【0067】

図 3 は、音源によって包絡形状 (正確には次数毎のケプストラム) の時間変動が異なる様子を示した図である。

(a) 非音声信号の包絡形状の時間変動、

(b) 音声信号の包絡形状の時間変動

これらの 2 つの異なる音源からの音声信号の包絡形状の時間変動の例を示している。

縦軸が振幅 (周波数)、横軸が時間に対応する。

【0068】

(a) の非音声信号の包絡形状の時間変動では、低周波から高周波まで一様な周期成分がランダムな位相を伴って混合されていることがうかがえる。

一方、(b) の音声信号の包絡形状の時間変動は、音の立ち上がりや立ち下がりが一定の周期 (主に 25 Hz 以下) を含んで規則的に変動している。

これらのことから、音声信号と非音声信号が混合された信号の場合、25 Hz 未満の時間変動に関しては音声信号が比較的支配的であり、25 Hz 以上の時間変動に関しては非音声信号が比較的支配的であると判断できる。

【0069】

従って、25 Hz 以上の高周波時間変動成分を除去、または低減することにより、非音声信号の時間変動を抑える効果があり、それと同時にフレーム間の急激な時間変動を抑え安定させる効果があると推定できる。

【0070】

図 4 は、音声信号の包絡形状に DC 成分が含まれる場合と、含まれない場合の包絡形状の時間変動を示している。

(c) DC 成分がない音声信号の包絡形状の時間変動

(d) DC 成分がある音声信号の包絡形状の時間変動

これらの 2 つの音声信号の包絡形状の時間変動の例を示している。

縦軸が振幅 (周波数)、横軸が時間に対応する。

【0071】

(c) に示す DC 成分を含まない音声信号の包絡形状の時間変動データは、全区間平均

10

20

30

40

50

を計算すると理論上平均値は 0 になる。

一方、(d) に示す DC 成分を含む音声信号の包絡形状の時間変動データは、全区間平均を計算すると、理論上の平均値が DC 成分に等しくなる。

このように計算される時間方向の DC 成分は、ケプストラム次数毎に異なる。

【0072】

図 5 には包絡形状 DC 成分の時系列の様子を示している。左奥から手前まで、ケプストラムの 1 次から R 次までを並べており、手前から右奥まで、ケプストラムの時間変動を並べている。

1 次から R 次までのケプストラム成分の各々は、時間変動をしており、各々、固有の DC 成分を持っている。

1 次から R 次までの DC 成分を周波数変換して、パワースペクトル軸に戻して観察した場合、時不変の周波数包絡の形を求めることができる。

【0073】

図 6 は、包絡形状の DC 成分の周波数領域の様子を示したものである。

図 6 (A) は、ケフレンシー領域で DC 成分として観察された 1 次から R 次までのケプストラムを示している。

この図 6 (A) に示すケフレンシー領域で DC 成分として観察された 1 次から R 次までのケプストラムを周波数変換してパワースペクトル領域に戻したデータが、図 6 (B) に示すデータである。

この図 6 (B) に示すように、静的な周波数特性が観察される。

このように、1 次から R 次までの DC 成分を周波数変換して、パワースペクトル軸に戻して観察することで、静的な周波数包絡の形を求めることができる。

【0074】

この図 6 (B) に示す DC 成分の周波数特性は時間変動に依存せず一定の周波数包絡であり、收音時のマイクのアナログ特性や残響成分に相当する場合や、あるいは、コーデックのプリ・ポストフィルタの特性に相当する場合であることが多い。

これらの DC 成分を除去することにより、乗法性歪み（マイク特性、残響）を低減する効果がある。

【0075】

以上の事実を考慮し、学習処理部 110 の包絡情報整形部 114 と、解析処理部 120 の包絡情報整形部 123 は、包絡情報整形処理におけるフィルタ通過帯域を多くの音源において音声の時間包絡が取りうる時間変動を考慮した処理として行うことが好ましい。

学習処理部 110 の包絡情報整形部 114 と、解析処理部 120 の包絡情報整形部 123 は、例えば以下の (式 5) に従って、整形包絡情報を生成する。

【0076】

【数 5】

$$C'_{low}(i, l) = \sum_{m=0}^{M_B-1} b(m) * C_{low}(i, l-m) + \sum_{m=1}^{M_A-1} a(m) * C'_{low}(i, l-m)$$

・・・ (式 5)

【0077】

なお、上記 (式 5) においては、変調周波数を 100 Hz (= 1 / (0.02 * 0.5)) と設定し、フィルタの伝達関数の分子の係数 b (m) を [0.25, 0.25, -0.25, -0.25]、分母の係数 a (m) を [1, -0.98] とし、それぞれ係数の総数を MB = 4、MA = 2 としている。

なお、係数 a (m), b (m) は、変調周波数に応じて設定可能である。

【0078】

学習処理部 110 の包絡情報整形部 114 は、サンプリング周波数 (f_{s2}) の学習用音声信号 51 に基づいて、

低域包絡算出部 112 が生成した低域包絡情報、すなわち、上記 (式 3) に従って算出した低域ケプストラム $C_{low}(i, 1)$

を入力し、包絡情報: $C_{low}(i, 1)$ に対して、上記 (式 5) に従って、整形包絡情報: $C'_{low}(i, 1)$ を生成して包絡情報学習部 116 に提供する。

【0079】

一方、解析処理部 120 の包絡情報整形部 123 は、サンプリング周波数 (f_{s1}) の入力音声信号 81 に基づいて、

低域包絡算出部 122 が生成した低域包絡情報、すなわち、上記 (式 3) に従って算出した低域ケプストラム $C_{low}(i, 1)$

を入力し、この包絡情報: $C_{low}(i, 1)$ に対して、上記 (式 5) に従って、整形低域包絡情報、すなわち整形低域ケプストラム情報 ($C'_{low}(i, 1)$) を生成して高域包絡形状推定部 125 に提供する。

【0080】

(2. 5. 包絡利得学習部と包絡形状学習部について)

包絡利得学習部 115 と包絡形状学習部 116 は、図 1 に示すように学習処理部 110 に設定されている。

【0081】

包絡利得学習部 115 と包絡形状学習部 116 は、学習用音声信号 51 に基づいて生成された包絡情報としての以下の情報、すなわち、

低域ケプストラム情報 $C_{low}(i, 1)$ 、

高域ケプストラム情報 $C_{high}(i, 1)$ 、

整形ケプストラム情報 $C'_{low}(i, 1)$ 、

これらの情報に基づいて、学習用音声信号 51 における低域と高域の包絡情報の関係を学習する。

具体的には、包絡利得学習部 115 は、低域の包絡利得情報からの高域の包絡利得情報を推定するための包絡利得情報として、[包絡利得推定情報 A] を算出する。

また、包絡形状学習部 116 は、低域の包絡形状情報からの高域の包絡形状情報を推定するための包絡形状情報として、[混合数 P]、[混合係数 π_p]、[平均 μ_p]、[共分散 Σ_p] を算出する。

【0082】

包絡利得学習部 115 と包絡形状学習部 116 は、包絡利得と包絡形状を別々に推定する。

包絡利得学習部 115 では、包絡利得をケプストラムの 0 次成分の推定処理として実行する。

包絡形状学習部 116 では、包絡形状をケプストラムの 0 次以外の低次の成分の推定によって実現している。

【0083】

具体的には、包絡利得学習部 115 では、例えば回帰式によってケプストラムの 0 次成分の推定処理を行い、包絡利得を算出する。

一方、包絡形状学習部 116 では、例えば、GMM (混合ガウスモデル) によってケプストラムの 0 次以外の低次の成分の推定を行い、包絡形状を算出する。

【0084】

包絡利得学習部 115 における包絡利得推定処理においては、説明変数として低域ケプストラム情報 $C_{low}(i, 1)$ の 0 次から R 次までの成分とそれぞれの 2 乗値を用い、被説明変数として高域ケプストラム情報の 0 次成分 $C_{high}(0, 1)$ を用いている。上記の説明変数の線形結合による推定値 (切片項も含む) と、目標値である被説明変数との間の二乗和誤差関数 $E(A)$ を最小にするような線形結合係数 A を [包絡利得推定情報 A] として求める。二乗和誤差関数 $E(A)$ は、以下の (式 6) によって示される。

【 0 0 8 5 】

【 数 6 】

$$\hat{C}_{high}(0, l) = A(0) + \sum_{i=0}^R A(i+1) * C_{low}(i, l) + \sum_{i=0}^R A(i+R+2) * (C_{low}(i, l))^2$$

$$E(A) = \frac{1}{2} \sum_{l=0}^{L-1} (\hat{C}_{high}(0, l) - C_{low}(0, l))^2$$

10

・ ・ ・ ・ ・ (式 6)

【 0 0 8 6 】

なお、上記（式 6）において、

例えば、一例として、 $R = 4$ とし、2 乗項を含めた非線形回帰を行う。

なお、その他の R の値の使用や、ニューラルネットワークやカーネル回帰などのその他の回帰手法を用いても構わない。

【 0 0 8 7 】

また、包絡形状学習部 1 1 6 における包絡形状の推定では、例えば GMM（ガウス混合モデル）を用いた処理を行う。

20

包絡形状学習部 1 1 6 における包絡形状の推定では、例えば GMM（ガウス混合モデル）を用いて、ケプストラムの 0 次以外の低次の成分の推定を行い、包絡形状を算出する。具体的には、包絡形状情報である「混合数 P 」、「混合係数 π_p 」、「平均 μ_p 」、「共分散 Σ_p 」を算出する。

【 0 0 8 8 】

なお、包絡形状の推定処理として行うケプストラムの 0 次以外の低次の成分の推定処理手法としては、GMM（ガウス混合モデル）を用いた処理の他、例えば、コーデックにおいてベクトル量子化の手段として頻繁に用いられる Kmeans 手法の適用も可能である。しかし、Kmeans と比較して、GMM は自由度の高いモデル化手法である。具体的には、例えば、包絡形状のクラスタリング手法（ベクトル量子化手法）を用いた処理が適用可能である。なお、GMM の全てのクラスタにおける共分散の自由度を減らし、単位行列にした場合、理論的には Kmeans とほぼ等しくなる。

30

【 0 0 8 9 】

図 7 ～ 図 9 は、Kmeans と GMM によるモデル化の比較を示す図である。

なお、図 7 ～ 図 9 に示すモデルは、多次元特徴量空間を 2 次元に簡略化して示している。

図 7 には、以下のモデル化データ例を示している。

(a) Kmeans（クラスタ数： $P = 1$ ）によってモデル化を行った例

(b) Kmeans（クラスタ数： $P > 1$ ）によってモデル化を行った例

(c) GMM（クラスタ数： $P = 1$ ）によってモデル化を行った例

(d) GMM（クラスタ数： $P > 1$ ）によってモデル化を行った例

40

【 0 0 9 0 】

図 7 (a) は、Kmeans（クラスタ数： $P = 1$ ）によってモデル化を行った例を示している。

図の円の外側を囲んでいる歪んだ形をした図形が、空間におけるデータ分布であるとした場合、Kmeans（クラスタ数： $P = 1$ ）によってモデル化を行うと超球状に分布のモデル化が行われ、表現しきれない箇所が多々出現する。図 7 において、グレーの円または楕円部分がモデル化されている空間であり、残りの箇所がモデル化できなかった空間である。

【 0 0 9 1 】

50

このようにKmeansのような超球状のモデルでは、1つのクラスタでは歪な空間を表現しきれないことが多い。よって、一般的には(b)のように、多数のクラスタを使用(クラスタ数： $P > 1$)し、空間の分布を埋めることが多い。

【0092】

一方、

(c) GMM (クラスタ数： $P = 1$) によってモデル化を行った例

の場合、モデルの共分散の自由度により、超球状から超楕円状に形を柔軟に変化させることが可能なため、データの分布に当てはまる体積がKmeansより大きくなる。

(d) GMM (クラスタ数： $P > 1$) によってモデル化を行った例

のように複数のクラスタを使用した場合においても、各クラスタの大きさ、方向、形を個別に変形させることができるため、分布への当てはまる体積が大きい。

10

【0093】

図7から理解される通り、

(a) Kmeans (クラスタ数： $P = 1$) によってモデル化を行った例

より、

(c) GMM (クラスタ数： $P = 1$) によってモデル化を行った例

の方が同じクラスタ数1では、データの分布をより正確に表現できている。

【0094】

また、(b)と(c)の比較においては、どちらも(a)よりも分布を正確に表現しているが、(b)の方が必要なクラスタ数が多く、その情報を保持するためのメモリーを必要とする。一方(c)に示すGMMでは、各クラスタの共分散情報を保持しており、その情報がクラスタの大きさ、方向、形を決定している。対角成分以外全て零という自由度の縛りを入れたモデル(対角共分散モデル)の場合、同じクラスタ数という条件のもとで、Kmeansの二倍のメモリーを必要とする。Kmeansがクラスタの平均値の情報のみを保持しているのに対し、GMMでは対角共分散の情報を保持していることに起因する。

20

【0095】

しかし、実際にはGMMによる表現能力は極めて高く、実施例のような音声の包絡形状をモデル化の対象とした場合、GMMに比較して、Kmeansは約四倍のクラスタ数を必要とするため、結局Kmeansのメモリーコストの方が大きい。また、計算量に関しても、Kmeansと比較した場合、クラスタ数と同数のlogの計算負荷が追加コストとなるが、FFTなどの計算負荷と比較すると非常に小さい値である。

30

【0096】

このような理由から、包絡形状学習部116における包絡形状の推定では、例えばGMM(ガウス混合モデル)を用いた処理を行う。

包絡形状学習部116における包絡形状の推定では、GMM(ガウス混合モデル)を用いて、ケプストラムの0次以外の低次の成分の推定を行い、包絡形状を算出する。具体的には、包絡形状情報である[混合数 P]、[混合係数 π_p]、[平均 μ_p]、[共分散 Σ_p]を算出する。

【0097】

40

実際の学習処理では、整形ケプストラム情報 $C'_{low}(i, l)$ および $C_{high}(i, l)$ を一つの結合ベクトル $C_{all}(i, l)$ とし、EMアルゴリズムによる対数事後確率最大化によって、 P 個のガウス分布のパラメータ、混合係数 π_p 、平均 μ_p 、共分散 Σ_p を求める。

具体的には、以下に示す(式7)に従って、包絡形状情報である[混合数 P]、[混合係数 π_p]、[平均 μ_p]、[共分散 Σ_p]を算出する。

【0098】

【数 7】

$$C_{all}(r-1, l) = \begin{cases} C'_{low}(r, l) * \alpha_{low}(r-1) & r = 1, \dots, R \\ C_{high}(r-R, l) * \alpha_{high}(r-R-1) & r = R+1, \dots, 2 * R \end{cases}$$

$$w_p(l) = \frac{\pi_p * N(C_{all}(l) | \mu_p, \Sigma_p)}{\sum_{n=0}^{P-1} \pi_n * N(C_{all}(l) | \mu_n, \Sigma_n)} \quad \left(\begin{array}{l} w_p : \text{各ガウス分布の負担率} \\ L_p : \text{各ガウス分布に属するフレーム数} \end{array} \right)$$

10

$$\mu_p^{new} = \frac{1}{L_p} \sum_{l=0}^{L-1} w_p(l) * C_{all}(l)$$

$$\Sigma_p^{new} = \frac{1}{L_p} \sum_{l=0}^{L-1} w_p(l) * (C_{all}(l) - \mu_p^{new}) * (C_{all}(l) - \mu_p^{new})^T$$

20

$$\pi_p^{new} = \frac{L_p}{L}$$

$$L_p^{new} = \sum_{l=0}^{L-1} w_p(l)$$

・・・・・・・・（式 7）

【0099】

30

なお、結合ベクトルを作成する際、整形ケプストラム情報 $C'_{low}(i, l)$ および $C_{high}(i, l)$ に、それぞれ予め決められた重み係数 $\alpha_{low}(r)$ 、 $\alpha_{high}(r)$ を乗じる。例えば、 $R=4$ とし、重み係数 $\alpha_{low}(r)$ 、 $\alpha_{high}(r)$ とともに $[0.5, 0.75, 1.0, 1.25]$ とする。なお、重み係数の設定は様々な設定が可能である。

【0100】

このように、

包絡利得学習部 115 では、

説明変数：低域ケプストラム情報 $C_{low}(i, l)$ の 0 次から R 次までの成分とそれぞれの 2 乗値、

40

被説明変数：高域ケプストラム情報の 0 次成分 $C_{high}(0, i)$ を用い、

説明変数の線形結合による推定値（切片項も含む）と、目標値である被説明変数との間の二乗和誤差関数 $E(A)$ を前記の（式 6）に従って算出し、この二乗和誤差関数 $E(A)$ を最小にするような線形結合係数 A を「包絡利得推定情報 A 」として求める。

【0101】

また、包絡形状学習部 116 では、上述したように例えば GMM（ガウス混合モデル）を用いて、ケプストラムの 0 次以外の低次の成分の推定を行い、包絡形状を算出する。具体的には、包絡形状情報である「混合数 P 」、「混合係数 π_p 」、「平均 μ_p 」、「共分散 Σ_p 」を算出する。

【0102】

50

図 1 に示すように、包絡利得学習部 115 の算出した [包絡利得推定情報 A] は、解析処理部 120 の高域包絡利得推定部 124 に提供される。

また、包絡形状学習部 116 の算出した包絡形状情報である [混合数 P]、[混合係数 π_p]、[平均 μ_p]、[共分散 Σ_p] は、解析処理部 120 の高域包絡形状推定部 125 へ提供される。

【0103】

(2.6. 高域包絡形状推定部について)

次に、図 1 に示す解析処理部 120 内に設けられた高域包絡形状推定部 125 の処理について説明する。

解析処理部 120 内の高域包絡形状推定部 125 は、解析処理部 120 内の包絡情報整形部 123 から入力音声信号 81 に基づいて生成された整形低域ケプストラム情報 $C'low(i, l)$ を入力する。

さらに、学習処理部 110 の包絡形状学習部 116 から、学習用音声信号 51 に基づく解析結果として得られた包絡形状情報である [混合数 P]、[混合係数 π_p]、[平均 μ_p]、[共分散 Σ_p] を入力する。

【0104】

高域包絡形状推定部 125 は、入力音声信号 81 に基づいて生成された整形低域ケプストラム情報 $C'low(i, l)$ に対して、学習用音声信号 51 に基づく包絡形状情報を利用した処理を実行することにより、入力音声信号 81 に対応する高域包絡形状情報 $C'high(i, l)$ の推定を行う。

ただし、 $i = 1, \dots, R$ である。

【0105】

図 8、図 9 を参照して、高域包絡形状推定部 125 の実行する高域包絡形状情報の推定処理について説明する。なお、先に説明したように、図 7～図 9 は、Kmeans と GMM によるモデル化の比較を示す図であり、図 7～図 9 に示すモデルは、多次元特徴量空間を 2 次元に簡略化して示している。

【0106】

図 8、図 9 は、Kmeans と GMM の二つの異なる手法を用いた場合、低域包絡形状 (写像元) から高域包絡形状 (写像先) への線形変換の様子が、どのように異なるかを描いた図である。

【0107】

Kmeans の場合、写像元がどのクラスタに属するかを、クラスタ重心までの距離を計算した後、属するクラスタの回帰直線を写像関数として、低域包絡形状から高域包絡形状への線形変換を行う。クラスタ重心と回帰係数は学習部において事前に決定されているものとする。

図 8 には、

(a) Kmeans + 線形回帰を利用した線形変換処理

(b) GMM の事後確率を利用した場合の線形変換処理

これらの処理例を説明する図である。

【0108】

図 8 (a) に示す Kmeans + 線形回帰を利用した線形変換処理例においては、二次元特徴量空間の分布に二つのクラスタ (クラスタ 1、クラスタ 2) が当てはめられている。学習処理時は写像元と写像先のデータが存在するため、双方のデータを用いてクラスタリングによる学習を行うが、帯域拡大処理時には写像先の情報は不明であり、写像元の低域包絡情報のみを保持しているため、写像元のデータのみを用いてクラスタ重心への距離計算を行い、クラスタリングを行う。

【0109】

図 8 (a) に示す例において、クラスタ重心までの距離がクラスタ 1 の方が小さい場合には回帰直線 1 を、クラスタ 2 の方が小さい場合には回帰直線 2 を使用して線形変換を行い写像先の結果を求める。仮にクラスタ境界にデータが存在する場合、写像関数が、回帰

直線 1 から回帰直線 2 へ切り替わるため、得られる結果が不安定であり、時間方向の不連続が起きやすい。

【0110】

図 8 (b) に示す GMM の事後確率を利用した場合の線形変換処理例においては、基本的には *Kmeans* と同様、距離を測り属するクラスタを求めるが、*Kmeans* と異なる点として、データが各クラスタにどの程度存在する可能性があるかを確率として計算することが可能であるという特徴がある。

【0111】

図 8 (b) に示す例では、クラスタ 1 に近いほど、クラスタ 1 に属する確率は高く、クラスタ 2 に属する確率は低くなる。クラスタ 2 に近い場合は逆の結果となる。この特徴を利用して、各クラスタにおけるデータの存在確率（一般的にはクラスの事後確率と呼ばれる）を用いて、各々のクラスタの回帰直線を混合することにより滑らかな混合曲線を作ることができ、連続的な写像を実現できる。図 8 の (b) では二つのクラスタの各々の存在確率を使用し、回帰直線を混合して混合曲線を描いたものである。写像元のデータはこの連続的な混合曲線によって写像される。

【0112】

図 9 は、図 8 と同様、

(a) *Kmeans* + 線形回帰を利用した線形変換処理

(b) GMM の事後確率を利用した場合の線形変換処理

これらの処理例を説明する図である。

図 9 は、(a) *Kmeans* と (b) GMM をそれぞれ使用した場合、写像元のデータがクラスタ境界を越えて変化すると、写像先のデータがどのように変化するかを説明する図である。

写像元のデータの値が a から $a + \delta$ へと極小さな変化をした場合の例について示している。

【0113】

(a) *Kmeans* + 線形回帰を利用した線形変換処理

この処理を行った場合は、図 9 (a) に示すように、クラスタが 1 から 2 へと変化するため、線形変換に使用される回帰係数が大きく変化し、写像先において大きな値の変化をする。

一方、

(b) GMM の事後確率を利用した場合の線形変換処理

この処理を行った場合は、図 9 (b) に示すように、クラスタ 1 からクラスタ 2 へと変化するが、与えられる写像関数は、存在確率によって混合され連続的な混合曲線であるため、写像先においても小さな値の変化しか起こらない。

この現象は、推定結果の時間方向の滑らかさ、という形で知覚される。

【0114】

このように、GMM を使用した手法においては、フレーム間にて滑らかな推定を行うことが可能であり、自然界に存在する音響信号の時間変動に比較的近い結果となる。仮にクラスタ間の距離が大きい場合、*Kmeans* による手法では音質的不連続を招くが、GMM による手法では連続的である。また、多くのクラスタを配置しなくても、クラスタの間を補間する効果も期待できるため、*Kmeans* と比較して少ないクラスタ数で実現でき、コストパフォーマンスの面で優れていると言える。

【0115】

このように、図 1 に示す解析処理部 120 内に設けられた高域包絡形状推定部 125 は、解析処理部 120 内の包絡情報整形部 123 から入力音声信号 81 に基づいて生成された整形低域ケプストラム情報 $C'_{low}(i, l)$ を入力し、学習処理部 110 の包絡形状学習部 116 から入力する学習用音声信号 51 に基づく解析結果として得られた包絡形状情報を利用して、GMM 手法を適用して、以下に示す (式 8) に従って、入力音声信号 81 に対応する高域包絡形状情報 $C_{high}(i, l)$ の推定を行う。

【 0 1 1 6 】

具体的には、GMM手法を適用した以下に示す（式 8）に従って、学習処理部 1 1 0 の包絡形状学習部 1 1 6 から入力する包絡形状情報である〔混合数 P〕、〔混合係数 π_p 〕、〔平均 μ_p 〕、〔共分散 Σ_p 〕を適用して、入力音声信号 8 1 に対応する高域包絡形状情報 $C^{high}(i, l)$ を算出する。

【 0 1 1 7 】

【 数 8 】

$$\hat{C}_{high}(l) = \sum_{p=0}^{P-1} w_p * \hat{y}_p \quad 10$$

$$C_{low}''(r-1, l) = C_{low}'(r, l) * \alpha_{low}(r-1) \quad r=1, \dots, R$$

$$w_p = \frac{\pi_p * N(C_{low}''(l) | \mu_p^{low}, \Sigma_p^{lowlow})}{\sum_{n=0}^{P-1} \pi_n * N(C_{low}''(l) | \mu_n^{low}, \Sigma_n^{lowlow})} \quad \left(\begin{array}{l} w_p : \text{各ガウス分布の負担率} \\ \hat{y}_p : \text{各ガウス分布における高域への写像結果} \\ \hat{C}_{high} : \text{推定高域ケプストラム} \end{array} \right)$$

$$\hat{y}_p = \mu_p^{high} + \Sigma_p^{highlow} (\Sigma_p^{lowlow})^{-1} (C_{low}''(l) - \mu_p^{low}) \quad 20$$

$$\mu_p = \begin{pmatrix} \mu_p^{low} \\ \mu_p^{high} \end{pmatrix}$$

$$\Sigma_p = \begin{pmatrix} \Sigma_p^{lowlow} & \Sigma_p^{lowhigh} \\ \Sigma_p^{highlow} & \Sigma_p^{highhigh} \end{pmatrix}$$

30

.....（式 8）

【 0 1 1 8 】

このように、高域包絡形状推定部 1 2 5 は、入力音声信号 8 1 に基づいて生成された整形低域ケプストラム情報 $C^{low}(i, l)$ に学習時と同じ重み係数 $\alpha_{low}(r)$ を乗じたのち、学習用音声信号 5 1 に基づく包絡形状情報を利用した処理により、入力音声信号 8 1 に対応する高域包絡形状情報 $C^{high}(i, l)$ の推定を行う。

ただし、 $i = 1, \dots, R$ である。

高域包絡形状推定部 1 2 5 は、上記（式 8）に従って算出した推定高域ケプストラム $C^{high}(i, l)$ を高域包絡補正部 1 2 7 へ供給する。

【 0 1 1 9 】

40

（ 2 . 7 . 高域包絡利得推定部について）

次に、図 1 に示す解析処理部 1 2 0 内に設けられた高域包絡利得推定部 1 2 4 の処理について説明する。

解析処理部 1 2 0 内の高域包絡利得推定部 1 2 4 は、解析処理部 1 2 0 内の低域包絡算出部 1 2 2 から入力音声信号 8 1 に基づいて生成された低域ケプストラム情報 $C^{low}(i, l)$ を入力する。

さらに、学習処理部 1 1 0 の包絡利得学習部 1 1 5 から、学習用音声信号 5 1 に基づく解析結果として得られた包絡利得情報である〔回帰係数 A〕を入力する。

【 0 1 2 0 】

高域包絡利得推定部 1 2 4 は、入力音声信号 8 1 に基づいて生成された低域ケプストラ

50

ム情報 $C_{low}(i, l)$ に対して、学習用音声信号 51 に基づく包絡利得情報である「回帰係数 A」を利用した処理を実行することにより、入力音声信号 81 に対応する高域包絡利得の推定を行う。

【0121】

具体的には、回帰モデルによって高域包絡利得の推定を行い、以下に示す（式 9）に従って、高域ケプストラムの 0 次成分 $\hat{C}_{high}(0, l)$ を推定する。ただし、 $i = 0, \dots, R$ である。

【0122】

【数 9】

$$\hat{C}_{high}(0, l) = A(0) + \sum_{i=0}^R A(i+1) * C_{low}(i, l) + \sum_{i=0}^R A(i+R+2) * (C_{low}(i, l))^2$$

・・・・・・（式 9）

【0123】

なお、高域ケプストラムの 0 次成分 $\hat{C}_{high}(0, l)$ は、高域包絡利得情報を表している。例えば、 $R = 4$ とし、2 乗項を含めた非線形回帰を行う。ただし高域包絡利得の推定処理としては、上記式に従った処理の他、ニューラルネットワークやカーネル回帰などのその他の回帰手法を用いてもよい。

【0124】

高域包絡利得推定部 124 が上記（式 9）に従って算出した高域包絡利得情報 $\hat{C}_{high}(0, l)$ は、高域包絡補正部 127 に供給される。

【0125】

（2. 8. 中域包絡補正部について）

次に、図 1 に示す解析処理部 120 内に設けられた中域包絡補正部 126 の処理について説明する。

解析処理部 120 内の中域包絡補正部 126 は、解析処理部 120 内の周波数分析部 121 から入力音声信号 81 に基づいて生成された時間周波数スペクトル $X_{ana}(k, l)$ を入力する。

さらに、解析処理部 120 内の低域包絡算出部 122 から、入力音声信号 81 に基づいて生成された低域ケプストラム $C_{low}(i, l)$ を入力する。

【0126】

中域包絡補正部 126 は、周波数分析部 121 から入力音声信号 81 に基づいて生成された時間周波数スペクトル $X_{ana}(k, l)$ の中域の周波数帯域部分、例えば $f_{s1}/4$ 以上 $f_{s1}/2$ 以下のスペクトルに相当する部分と、低域包絡算出部 122 より供給された低域ケプストラム $C_{low}(i, l)$ を用いて、周波数軸上で平坦化されたスペクトル信号を生成する。

【0127】

まず、低域ケプストラム $C_{low}(i, l)$ のうち、ケプストラムの低次数以外の係数を 0 にした後、パワースペクトル領域へ戻し、以下の（式 10）に従って、リフター低域スペクトル $X_{lfft_l}(k, l)$ を得る。

【0128】

【数 10】

$$X_{lfft_l}(k, l) = \exp\left(\sum_{i=0}^{M-1} C_{low}(i, l) * \exp\left(-j2\pi \frac{i * k}{M}\right)\right)$$

・・・・・・（式 10）

【 0 1 2 9 】

次に、中域包絡補正部 1 2 6 は、上記（式 1 0）に従って得られたリフター低域スペクトル $X_{lift_l}(k)$ の、中域部分（ $f_{s1}/4$ 以上 $f_{s1}/2$ 以下）のスペクトルに相当する部分（この場合、 $k = M/4, \dots, M/2$ ）を用いて、時間周波数スペクトル $X_{ana}(k, l)$ の同じ周波数部分を割り、平坦化を行った後、周波数 $f_{s1}/4$ を中心にして、低域側にミラーリングを行い、中域スペクトル $X_{white}(k, l)$ を得る。

中域スペクトル $X_{white}(k, l)$ は、以下に示す（式 1 1）に従って算出する。

【 0 1 3 0 】

【 数 1 1 】

10

$$X_{mid}(k, l) = \frac{X_{ana}(k, l)}{X_{lift_l}(k, l)} \quad k = \frac{M}{4}, \dots, \frac{M}{2}$$

$$X_{white}(k, l) = \begin{cases} \text{conj}\left(X_{mid}\left(\frac{M}{2} - k, l\right)\right) & k = 0, \dots, \frac{M}{4} - 1 \\ X_{mid}(k, l) & k = \frac{M}{4}, \dots, \frac{M}{2} \end{cases}$$

20

．．．．．（式 1 1）

【 0 1 3 1 】

中域包絡補正部 1 2 6 が、上記の（式 1 0）、（式 1 1）に従って算出した中域スペクトル $X_{white}(k, l)$ は、高域包絡補正部 1 2 7 へ供給される。

【 0 1 3 2 】

（ 2 . 9 . 高域包絡補正部について）

30

次に、図 1 に示す解析処理部 1 2 0 内に設けられた高域包絡補正部 1 2 7 の処理について説明する。

解析処理部 1 2 0 内の高域包絡補正部 1 2 7 は、解析処理部 1 2 0 内の中域包絡補正部 1 2 6 から入力音声信号 8 1 に基づいて生成された中域スペクトル $X_{white}(k, l)$ を入力する。

さらに、解析処理部 1 2 0 内の高域包絡利得推定部 1 2 4 から、学習データである包絡利得情報を利用して推定された入力音声信号 8 1 の高域包絡利得情報 $Chigh(0, l)$ を入力する。

さらに、解析処理部 1 2 0 内の高域包絡形状推定部 1 2 5 から、学習データである包絡形状情報を利用して推定された入力音声信号 8 1 の高域包絡形状情報 $Chigh(i, h)$ を入力する。

40

【 0 1 3 3 】

高域包絡補正部 1 2 7 は、これらの入力情報に基づいて、入力音声信号 8 1 の高域包絡情報の補正を行う。具体的な処理は以下の通りである。

【 0 1 3 4 】

高域包絡補正部 1 2 7 は、中域包絡補正部 1 2 6 から入力音声信号 8 1 に基づいて生成された中域スペクトル $X_{white}(k, l)$ を入力し、この中域スペクトル $X_{white}(k, l)$ に対して、高域包絡利得推定部 1 2 4 の生成した高域包絡利得情報 $Chigh(0, l)$ と、高域包絡形状推定部 1 2 5 の生成した高域包絡利得情報 $Chigh(i, l)$ （ただし、 $i = 1, \dots, R$ ）を用いて、包絡を補正する。

50

【 0 1 3 5 】

まず、高域包絡利得推定部 1 2 4 の生成した高域包絡利得情報 $C_{high}(0, l)$ と高域包絡形状推定部 1 2 5 の生成した高域包絡利得情報 $C_{high}(i, l)$ をパワースペクトルによる包絡情報へ戻し、以下の (式 1 2) に従って、リフター高域スペクトル $X_{lift_h}(k, l)$ を得る。

【 0 1 3 6 】

【 数 1 2 】

$$X_{lift_h}(k, l) = \exp \left(\sum_{i=0}^{M-1} \hat{C}_{high}(i, l) * \exp \left(-j2\pi \frac{i * k}{M} \right) \right) \quad 10$$

・・・・・・・・ (式 1 2)

【 0 1 3 7 】

高域包絡補正部 1 2 7 は、上記 (式 1 2) に従って得られたリフター高域スペクトル $X_{lift_h}(k, l)$ を適用して、以下に示す (式 1 3) に従って、中域スペクトル $X_{white}(k, l)$ を補正し、補正済み中域スペクトル $X'_{white}(k, l)$ を得る。

20

【 0 1 3 8 】

【 数 1 3 】

$$X'_{white}(k, l) = X_{white}(k, l) * X_{lift_h}(k, l)$$

・・・・・・・・ (式 1 3)

【 0 1 3 9 】

30

さらに、高域包絡補正部 1 2 7 は、上記 (式 1 2) に従って補正されたスペクトル $X'_{white}(k, l)$ を周波数 $f_{s1}/2$ (この場合、 $k = M/2$) を中心にして反転させ、元々スペクトルが存在していた低域スペクトルには 0 を挿入し、以下の (式 1 4) に示す高域スペクトル $X_{high}(k, l)$ を得る。

【 0 1 4 0 】

【 数 1 4 】

$$X_{high}(k, l) = \begin{cases} 0 & k = 0, \dots, \frac{M}{2} - 1 \\ \text{conj}(X'_{white}(M - k, l)) & k = \frac{M}{2}, \dots, M - 1 \\ X'_{white}(k, l) & k = M, \dots, \frac{3M}{2} \\ 0 & k = \frac{3M}{2} + 1, \dots, 2M - 1 \end{cases} \quad 40$$

50

．．．．．（式１４）

【０１４１】

この結果、周波数 f_{s2} （この場合、FFTポイント数は $2M$ ）の高域スペクトル $X_{high}(k, l)$ 信号が生成される。

高域包絡補正部１２７の生成した高域スペクトル $X_{high}(k, l)$ は、周波数合成部１２８へ供給される。

【０１４２】

（２．１０．周波数合成部について）

次に、図１に示す解析処理部１２０内に設けられた周波数合成部１２８の処理について説明する。

10

周波数合成部１２８は、解析処理部１２０内の高域包絡補正部１２７から高域スペクトル $X_{high}(k, l)$ を入力する。

さらに、周波数分析部１２１から、入力音声信号８１に基づいて生成された周波数スペクトル $X_{ana}(k, l)$ を入力する。

【０１４３】

周波数合成部１２８は、解析処理部１２０内の高域包絡補正部１２７から高域スペクトル $X_{high}(k, l)$ と、周波数分析部１２１より供給された 0 以上 $f_{s1}/2$ 以下に相当する周波数スペクトル $X_{ana}(k, l)$ （この場合、 $k = 0, \dots, M/2$ ）の部分を用いて、以下の（式１５）に従って、合成スペクトル $X_{syn}(k, l)$ を求める。

【０１４４】

20

【数１５】

$$X_{syn}(k, l) = \begin{cases} X_{ana}(k, l) & k = 0, \dots, \frac{M}{2} - 1 \\ \frac{X_{ana}(k, l) + X_{high}(k, l)}{2} & k = \frac{M}{2} \\ X_{high}(k, l) & k = \frac{M}{2} + 1, \dots, \frac{3M}{2} - 1 \\ \frac{X_{ana}(2M - k, l) + X_{high}(k, l)}{2} & k = \frac{3M}{2} \\ \text{conj}(X_{ana}(2M - k, l)) & k = \frac{3M}{2} + 1, \dots, 2M - 1 \end{cases}$$

30

．．．．．（式１５）

【０１４５】

周波数合成部１２８は、上記（式１５）に従って算出した合成スペクトル $X_{syn}(k, l)$ に逆周波数変換を行い、時間領域の合成信号 $x_{syn}(n, l)$ を得る。

40

以下に示す（式１６）に従って、時間領域の合成信号 $x_{syn}(n, l)$ を得る。

【０１４６】

【数 1 6】

$$x_{syn}(n, l) = \frac{1}{M} \sum_{k=0}^{M-1} X_{syn}(k, l) * \exp\left(j2\pi n \frac{k}{M}\right)$$

・・・・・・・・ (式 1 6)

10

【0 1 4 7】

なお、本実施例においては、逆周波数変換に I D F T (逆離散フーリエ変換) を用いているが、周波数分析部で使用した変換の逆変換に相当するものを用いればよい。ただし、フレームの大きさ N は拡張された周波数 f s 2 における 0. 0 2 s e c 相当のサンプル数 (N = サンプリグ周波数 f s 2 * 0. 0 2) であり、D F T ポイント数 M は N 以上の 2 のべき乗の値であるため、上述に用いていた N、M とはサイズが異なることに注意が必要である。

【0 1 4 8】

周波数合成部 1 2 8 は、上記 (式 1 6) に従って算出した合成信号 x s y n (n , l) に窓関数 w _ s y n (n) を掛け、オーバーラップ加算を行うことで、フレーム合成を行い、出力信号 y (n) を生成する。。

20

【0 1 4 9】

具体的な出力信号 y (n) の算出式と、窓関数 w _ s y n (n) を以下の (式 1 7) に示す。

【0 1 5 0】

【数 1 7】

$$y(n + l * N) = x_{syn}(n, l) * w_{syn}(n) + y(n + l * N)$$

30

$$w_{syn}(n) = \begin{cases} \left(0.5 - 0.5 * \cos\left(2\pi \frac{n}{N}\right)\right)^{0.5} & n = 0, \dots, N - 1 \\ 0 & n = N, \dots, M - 1 \end{cases}$$

・・・・・・・・ (式 1 7)

【0 1 5 1】

40

なお、上記処理では、窓関数にハニング窓の平方根を用いて 5 0 % のオーバーラップ加算を行っているが、サイン窓などのその他の窓や、5 0 % 以外のオーバーラップ比率であっても構わない。

周波数合成部 1 2 8 が上記 (式 1 7) に従って算出した信号 y (n) が、図 1 に示す音声信号処理装置 1 0 0 の出力音声信号 8 2 として出力される。

この出力音声信号 8 2 はサンプリグ周波数 (f s 2) となり、入力音声信号のサンプリグ周波数 (f s 1) の 2 倍のサンナプリグ周波数を持つ周波数帯域の拡大された音声信号となる。

【0 1 5 2】

なお、上記の実施例では、図 1 に示す音声信号処理装置 1 0 0 として、学習処理部 1 1

50

0と、解析処理部120の2つの処理部を有する構成例として説明したが、学習処理部110の学習結果として得られる学習データを予め記憶部に格納する構成としてもよい。すなわち、記憶部に格納された学習データを解析処理部120が必要に応じて取得して入力信号に対する処理を行う構成としてもよい。この構成の場合は、学習処理部を省略した解析処理部と学習結果としての学習データを格納した記憶部によって音声信号処理装置を構成することか可能である。

【0153】

以上、特定の実施例を参照しながら、本発明について詳解してきた。しかしながら、本発明の要旨を逸脱しない範囲で当業者が実施例の修正や代用を成し得ることは自明である。すなわち、例示という形態で本発明を開示してきたのであり、限定的に解釈されるべきではない。本発明の要旨を判断するためには、特許請求の範囲の欄を参酌すべきである。

【0154】

また、明細書中において説明した一連の処理はハードウェア、またはソフトウェア、あるいは両者の複合構成によって実行することが可能である。ソフトウェアによる処理を実行する場合は、処理シーケンスを記録したプログラムを、専用のハードウェアに組み込まれたコンピュータ内のメモリにインストールして実行させるか、あるいは、各種処理が実行可能な汎用コンピュータにプログラムをインストールして実行させることが可能である。例えば、プログラムは記録媒体に予め記録しておくことができる。記録媒体からコンピュータにインストールする他、LAN (Local Area Network)、インターネットといったネットワークを介してプログラムを受信し、内蔵するハードディスク等の記録媒体にインストールすることができる。

【0155】

なお、明細書に記載された各種の処理は、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にあるいは個別に実行されてもよい。また、本明細書においてシステムとは、複数の装置の論理的集合構成であり、各構成の装置が同一筐体内にあるものには限らない。

【産業上の利用可能性】

【0156】

以上、説明したように、本発明の一実施例の構成によれば、音声信号に対する高精度な周波数帯域拡大処理を実行する装置、方法が実現される。

例えば、本発明の一実施例の構成においては、入力音声信号の周波数分析結果に基づいて、低周波数帯域の包絡情報である低域包絡情報を算出する。さらに、学習用音声信号に基づく学習データ、例えば低域包絡情報から高周波数帯域の包絡情報である高域包絡情報を算出する学習データを適用して、入力音声信号対応の低域包絡情報から入力信号対応の高域包絡情報を推定して生成する。さらに、推定処理により生成した入力信号対応の高域包絡情報に対応する高周波数帯域信号を入力音声信号に合成して周波数帯域を拡大した出力音声信号を生成する。高域の包絡利得と包絡形状を、学習データを用いて推定することで高精度な帯域拡大が実現される。

【符号の説明】

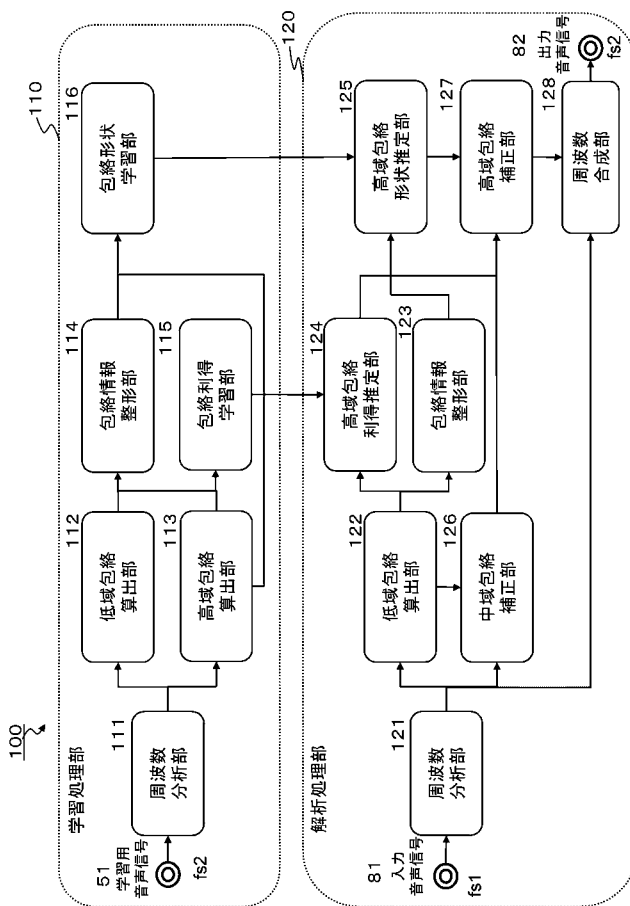
【0157】

- 51 学習用音声信号
- 81 入力音声信号
- 82 出力音声信号
- 100 音声信号処理装置
- 110 学習処理部
- 111 周波数分析部
- 112 低域包絡算出部
- 113 高域包絡算出部
- 114 包絡情報整形部
- 115 包絡利得学習部

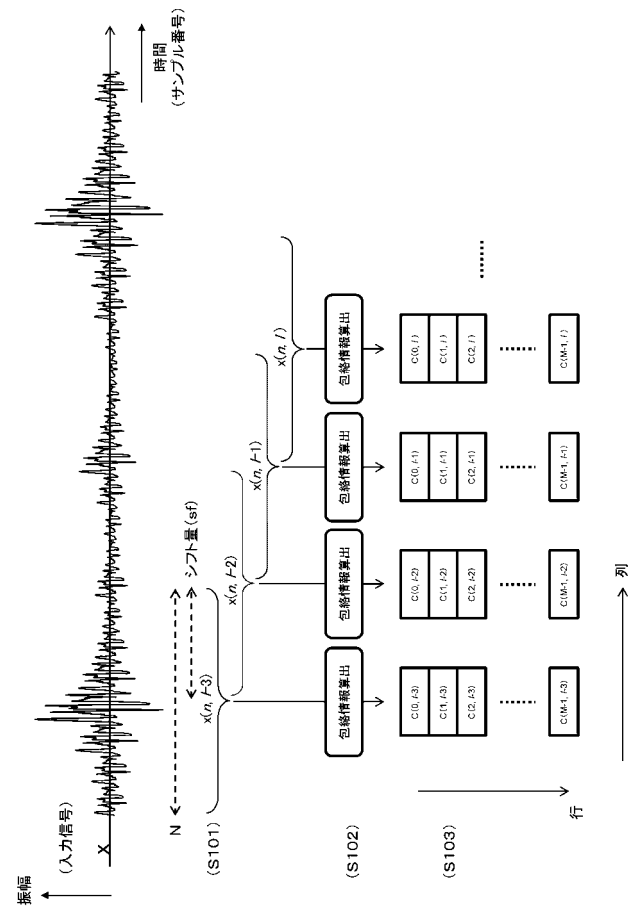
- 1 1 6 包絡形状学習部
- 1 2 0 解析処理部
- 1 2 1 周波数分析部
- 1 2 2 低域包絡算出部
- 1 2 3 包絡情報整形部
- 1 2 4 高域包絡利得推定部
- 1 2 5 高域包絡形状推定部
- 1 2 6 中域包絡補正部
- 1 2 7 高域包絡補正部
- 1 2 8 周波数合成部

10

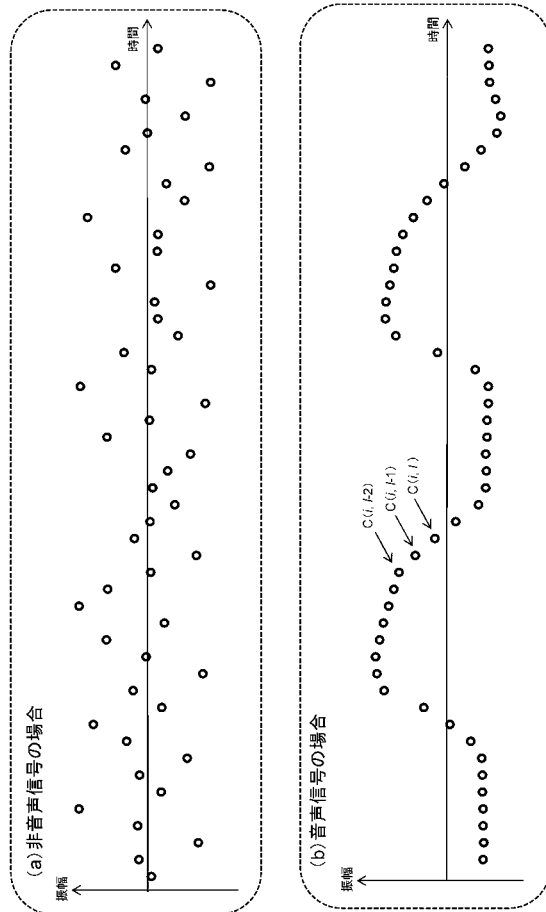
【図 1】



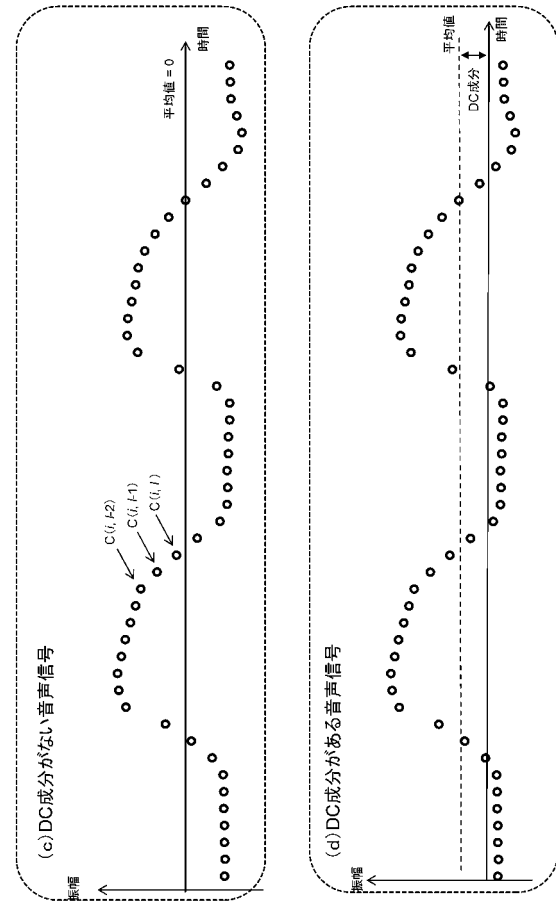
【図 2】



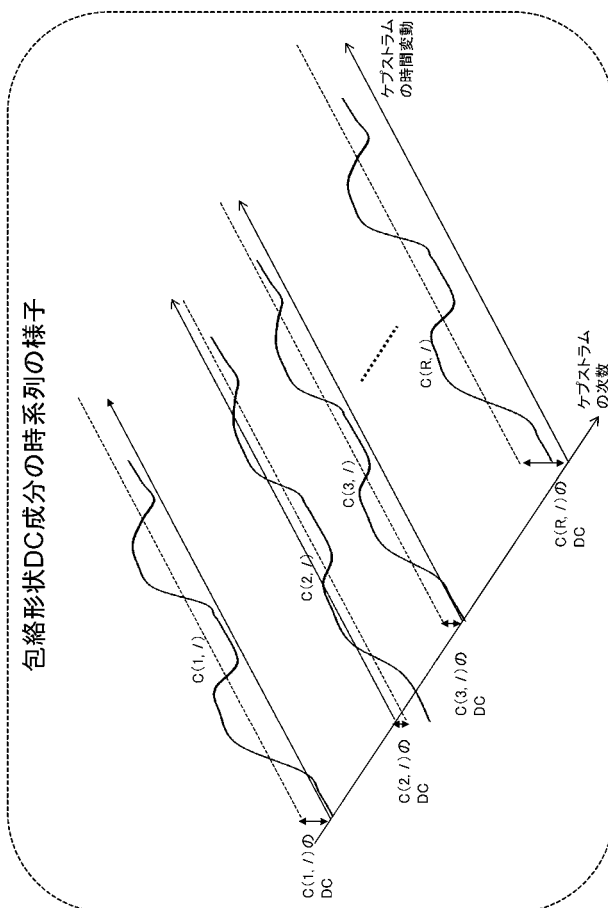
【図 3】



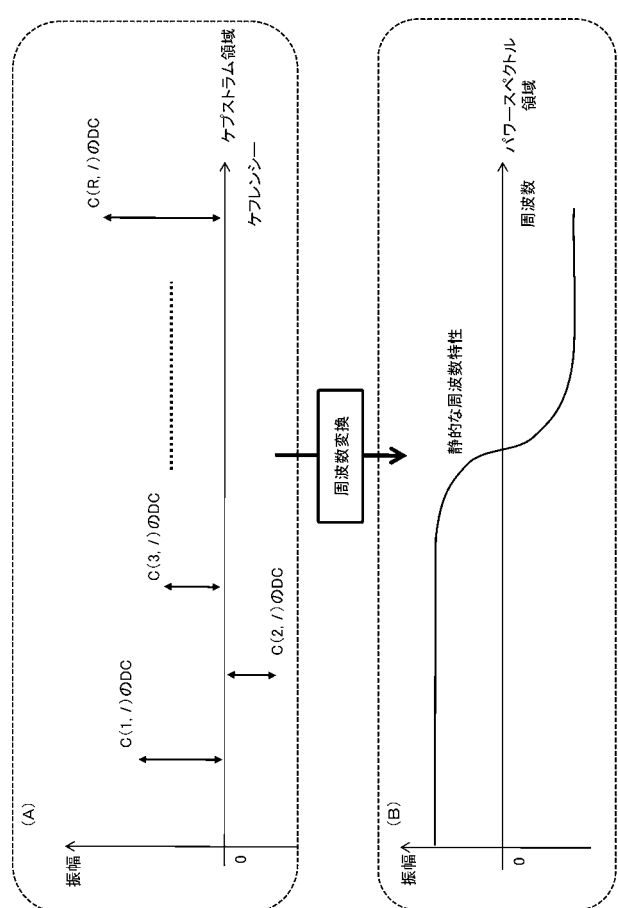
【図 4】



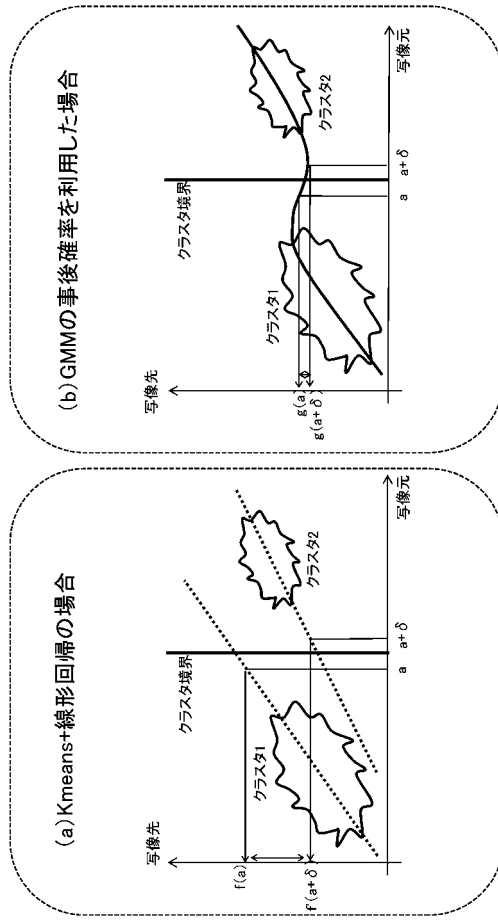
【図 5】



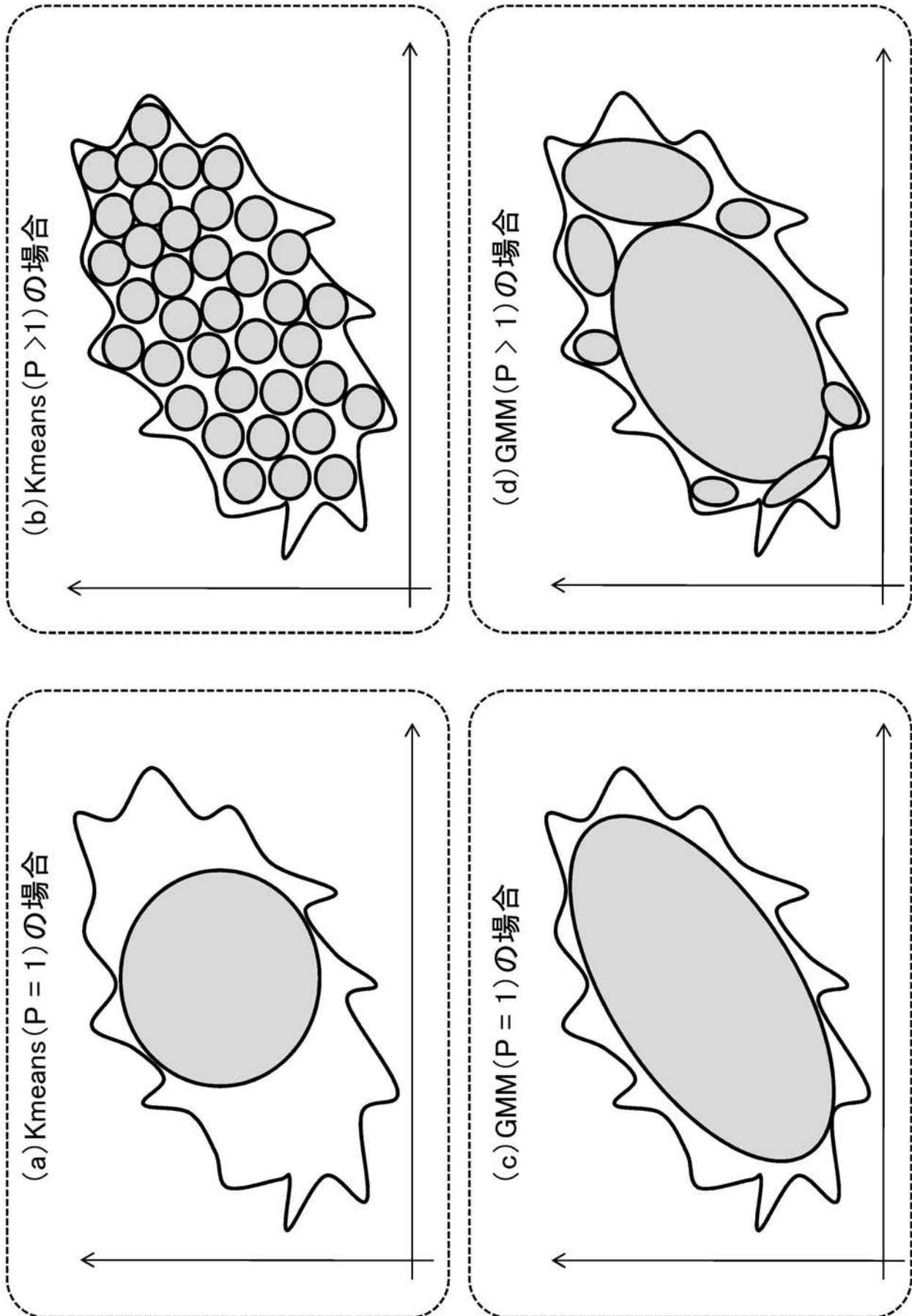
【図 6】



【図 9】



【図 7】



【図 8】

