(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2014-228779
(P2014-228779A)
(43)公開日　平成26年12月8日(2014.12.8)

(51)Int.Cl.		F I		テーマコード（参考）
G10L 19/005	(2013.01)	G10L 19/00	330E	
G10L 21/0388	(2013.01)	G10L 21/04	130A	

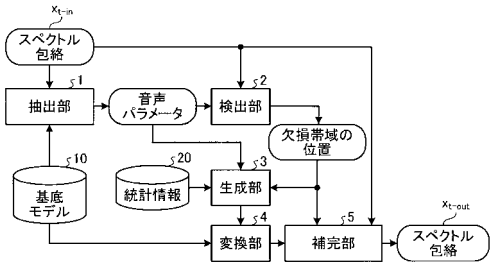
審査請求 未請求　請求項の数 11　OL　（全 14 頁）

(21)出願番号　特願2013-109897 (P2013-109897)
(22)出願日　平成25年5月24日 (2013.5.24)

(71)出願人　000003078
株式会社東芝
東京都港区芝浦一丁目１番１号
(74)代理人　100089118
弁理士　酒井　宏明
(74)代理人　100112656
弁理士　宮田　英毅
(72)発明者　大谷　大和
東京都港区芝浦一丁目１番１号　株式会社東芝内
(72)発明者　森田　眞弘
東京都港区芝浦一丁目１番１号　株式会社東芝内

(54)【発明の名称】音声処理装置、方法およびプログラム

(57)【要約】
【課題】任意の周波数帯域で欠損した音声成分を適切に補完することができる音声処理装置、方法およびプログラムを提供する。
【解決手段】実施形態の音声処理装置は、抽出部と、検出部と、生成部と、変換部と、補完部とを備える。抽出部は、入力音声のスペクトル包絡から、細分化された周波数帯域ごとの音声成分を表現する音声パラメータを抽出する。検出部は、入力音声のスペクトル包絡において音声成分が欠損している周波数帯域である欠損帯域を検出する。生成部は、欠損帯域の位置と、統計情報と、入力音声のスペクトル包絡から抽出された音声パラメータとに基づいて、欠損帯域に対応する音声パラメータを生成する。変換部は、欠損帯域に対応する音声パラメータを、欠損帯域のスペクトル包絡に変換する。補完部は、欠損帯域のスペクトル包絡と入力音声のスペクトル包絡とを合成して、欠損帯域が補完されたスペクトル包絡を生成する。
【選択図】図１



【特許請求の範囲】
【請求項１】
入力音声のスペクトル包絡から、細分化された周波数帯域ごとの音声成分を表現する音声パラメータを抽出する抽出部と、
前記入力音声のスペクトル包絡において音声成分が欠損している周波数帯域である欠損帯域を検出する検出部と、
検出された前記欠損帯域の位置と、音声成分が欠損していない音声のスペクトル包絡から抽出された前記音声パラメータを用いて事前に作成された統計情報と、前記入力音声のスペクトル包絡から抽出された前記音声パラメータとに基づいて、前記欠損帯域に対応する前記音声パラメータを生成する生成部と、
生成された前記欠損帯域に対応する前記音声パラメータを、前記欠損帯域のスペクトル包絡に変換する変換部と、
前記欠損帯域のスペクトル包絡と前記入力音声のスペクトル包絡とを合成して、前記欠損帯域が補完されたスペクトル包絡を生成する補完部と、を備える音声処理装置。
【請求項２】
前記音声パラメータは、細分化された前記周波数帯域の各々に対応する複数の基底ベクトルを用いて算出される値であり、
前記基底ベクトルの数は、音声のスペクトル包絡の分析に用いた分析点数よりも少ないことを特徴とする請求項１に記載の音声処理装置。
【請求項３】
前記基底ベクトルに対応する前記周波数帯域の範囲は、周波数軸上で隣り合う範囲の一部が重複していることを特徴とする請求項２に記載の音声処理装置。
【請求項４】
前記音声パラメータは、複数の前記基底ベクトルと各基底ベクトルに対応する重みベクトルとの線形結合と、音声のスペクトル包絡と、の誤差が最小になるように決定された前記重みベクトルであることを特徴とする請求項２または３に記載の音声処理装置。
【請求項５】
前記検出部は、前記入力音声のスペクトル包絡または該スペクトル包絡から抽出された前記音声パラメータの包絡形状を解析して、前記欠損帯域を検出することを特徴とする請求項１に記載の音声処理装置。
【請求項６】
前記統計情報は、音声成分が欠損していない複数の話者の音声から抽出された前記音声パラメータを学習データとして構築された統計モデルであることを特徴とする請求項１に記載の音声処理装置。
【請求項７】
前記統計情報は、音声成分が欠損していない複数の話者の音声から抽出された前記音声パラメータの系列と、該音声パラメータの系列から抽出された時間変動成分と、を学習データとして構築された統計モデルであることを特徴とする請求項１に記載の音声処理装置。
【請求項８】
前記生成部は、前記欠損帯域の位置と前記統計情報とに基づいて、前記欠損帯域を除く周波数帯域である残存帯域に対応する前記音声パラメータから前記欠損帯域に対応する前記音声パラメータを生成する規則を構築し、該規則を用いて、前記入力音声の音声スペクトル包絡から抽出された前記音声パラメータから、前記欠損帯域に対応する前記音声パラメータを生成することを特徴とする請求項１に記載の音声処理装置。
【請求項９】
前記変換部は、前記欠損帯域に対応する前記音声パラメータとして生成された前記重みベクトルと、前記欠損帯域に対応する前記基底ベクトルとを線形結合することにより、前記欠損帯域に対応する前記音声パラメータを前記欠損帯域のスペクトル包絡に変換することを特徴とする請求項４に記載の音声処理装置。

【請求項１０】
音声処理装置において実行される音声処理方法であって、
前記音声処理装置が、入力音声のスペクトル包絡から、細分化された周波数帯域ごとの音声成分を表現する音声パラメータを抽出するステップと、
前記音声処理装置が、前記入力音声のスペクトル包絡において音声成分が欠損している周波数帯域である欠損帯域を検出するステップと、
前記音声処理装置が、検出された前記欠損帯域の位置と、音声成分が欠損していない音声のスペクトル包絡から抽出された前記音声パラメータを用いて事前に作成された統計情報と、前記入力音声のスペクトル包絡から抽出された前記音声パラメータとに基づいて、前記欠損帯域に対応する前記音声パラメータを生成するステップと、
前記音声処理装置が、生成された前記欠損帯域に対応する前記音声パラメータを、前記欠損帯域のスペクトル包絡に変換するステップと、
前記音声処理装置が、前記欠損帯域のスペクトル包絡と前記入力音声のスペクトル包絡とを合成して、前記欠損帯域が補完されたスペクトル包絡を生成するステップと、を含む音声処理方法。
【請求項１１】
コンピュータに、
入力音声のスペクトル包絡から、細分化された周波数帯域ごとの音声成分を表現する音声パラメータを抽出する機能と、
前記入力音声のスペクトル包絡において音声成分が欠損している周波数帯域である欠損帯域を検出する機能と、
検出された前記欠損帯域の位置と、音声成分が欠損していない音声のスペクトル包絡から抽出された前記音声パラメータを用いて事前に作成された統計情報と、前記入力音声のスペクトル包絡から抽出された前記音声パラメータとに基づいて、前記欠損帯域に対応する前記音声パラメータを生成する機能と、
生成された前記欠損帯域に対応する前記音声パラメータを、前記欠損帯域のスペクトル包絡に変換する機能と、
前記欠損帯域のスペクトル包絡と前記入力音声のスペクトル包絡とを合成して、前記欠損帯域が補完されたスペクトル包絡を生成する機能と、を実現させるためのプログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明の実施の形態は、音声処理装置、方法およびプログラムに関する。
【背景技術】
【０００２】
従来、例えば携帯電話機や音声収録装置等の音声品質を向上させる技術として、帯域拡張が知られている。帯域拡張は、狭帯域音声から広帯域音声を構築する技術であり、例えば、入力音声において欠損している高周波帯域の音声成分を、欠損していない音声成分を用いて補完することができる。
【０００３】
しかし、従来の帯域拡張では、入力音声において欠損している高周波帯域の音声成分や、予め定められた特定の周波数帯域の音声成分を補完することはできるが、任意の周波数帯域の音声成分が部分的に欠損した場合に対応できない。音声処理装置に入力される音声信号は、伝送路の静的特性等の何らかの影響によって、任意の周波数帯域の音声成分が部分的に欠損することがあり、任意の周波数帯域の音声成分を適切に補完できるようにすることが求められる。
【先行技術文献】
【特許文献】
【０００４】
【特許文献１】特開２０１２－８３７９０号公報

【発明の概要】
【発明が解決しようとする課題】
【０００５】
本発明が解決しようとする課題は、任意の周波数帯域で欠損した音声成分を適切に補完することができる音声処理装置、方法およびプログラムを提供することである。
【課題を解決するための手段】
【０００６】
実施形態の音声処理装置は、抽出部と、検出部と、生成部と、変換部と、補完部と、を備える。抽出部は、入力音声のスペクトル包絡から、細分化された周波数帯域ごとの音声成分を表現する音声パラメータを抽出する。検出部は、前記入力音声のスペクトル包絡において音声成分が欠損している周波数帯域である欠損帯域を検出する。生成部は、検出された前記欠損帯域の位置と、音声成分が欠損していない音声のスペクトル包絡から抽出された前記音声パラメータを用いて事前に作成された統計情報と、前記入力音声のスペクトル包絡から抽出された前記音声パラメータとに基づいて、前記欠損帯域に対応する前記音声パラメータを生成する。変換部は、生成された前記欠損帯域に対応する前記音声パラメータを、前記欠損帯域のスペクトル包絡に変換する。補完部は、前記欠損帯域のスペクトル包絡と前記入力音声のスペクトル包絡とを合成して、前記欠損帯域が補完されたスペクトル包絡を生成する。
【図面の簡単な説明】
【０００７】
【図１】実施形態の音声処理装置の構成を示すブロック図。
【図２】実施形態の音声処理装置が実行する処理の流れを示すフローチャート。
【図３】検出部による欠損帯域の検出方法の一例を示す図。
【図４】補完部による処理の一例を示す図。
【図５】補完部による処理の他の例を示す図。
【発明を実施するための形態】
【０００８】

本実施形態の音声処理装置は、任意の周波数帯域の音声成分が欠損している入力音声のスペクトル包絡から、欠損している成分を補完したスペクトル包絡を生成する。入力音声は、主に、人の発話音声を想定している。図１は、実施形態の音声処理装置の構成を示すブロック図である。図２は、実施形態の音声処理装置が実行する処理の流れを示すフローチャートである。
【０００９】
本実施形態の音声処理装置は、図１に示すように、抽出部１と、検出部２と、生成部３と、変換部４と、補完部５と、を備える。
【００１０】
抽出部１は、入力音声のスペクトル包絡χｔ＿ｉｎから、基底モデル１０を用いて、細分化された周波数帯域ごとの音声成分を表現する音声パラメータを抽出する（図２のステップＳ１０１）。なお、入力音声からスペクトル包絡χｔ＿ｉｎを生成する処理は、音声処理装置の内部で行ってもよいし、外部で行ってもよい。
【００１１】
基底モデル１０は、音声のスペクトル包絡χｔによって形成される空間の部分空間の基底を表す基底ベクトルのセットである。本実施形態では、基底モデル１０として、下記の参考文献１に記載されたサブバンド基底スペクトルモデル（以下、ＳＢＭという。）を用いる。基底モデル１０は、音声処理装置内の図示しない記憶部に予め格納されてもよいし、音声処理装置の動作時に外部から取得されて保持されてもよい。
参考文献１：M Tamura, T Kagoshima, and M Akamine, “Sub-band basis spectrum model for pitch-synchronous log-spectrum and phase based on approximation of sparse coding, ”in Proceeding Interspeech 2010, pp. 2046－2049

, Sept. 2010.

【００１２】

参考文献１によれば、ＳＢＭの基底は、以下の（１）～（３）に示す特徴を持つ。

（１）周波数軸上で単一の最大値を与えるピーク周波数を含む所定の周波数帯域に値が存在し、その周波数帯域の外側は値を零とし、フーリエ変換やコサイン変換で用いられるような周期的な基底のように同じ最大値を複数持たない。

（２）基底の数は、スペクトル包絡がもつ分析点数よりも少なく、その数は分析点数の半分未満の数となる。

（３）ピーク周波数位置が隣りあう２つの基底間に重なりを持つ、すなわちピーク周波数が隣り合う基底は、値の存在する周波数の範囲の一部が重なる。

【００１３】

また、参考文献１によれば、ＳＢＭの基底を表す基底ベクトルは、下記式（１）により定義される。

【数１】

$$\phi_n(k)=\begin{cases}0.5-0.5\cos\left(\dfrac{k-\tilde{\Omega}(n-1)}{\tilde{\Omega}(n)-\tilde{\Omega}(n-1)}\pi\right) & \left(\tilde{\Omega}(n-1)\le k<\tilde{\Omega}(n)\right)\\ 0.5-0.5\cos\left(\dfrac{k-\tilde{\Omega}(n)}{\tilde{\Omega}(n+1)-\tilde{\Omega}(n)}\pi+\dfrac{\pi}{2}\right) & \left(\tilde{\Omega}(n)\le k<\tilde{\Omega}(n+1)\right)\\ 0 & (\text{otherwise})\end{cases} \quad \cdots (1)$$

ここで、Φｎ（ｋ）はｎ番目の基底ベクトルのｋ番目の成分である。また、Ω（ｎ）［ｒａｄ］はｎ番目の基底ベクトルのピーク周波数であり、下記式（２）のように定義される。

【数２】

$$\tilde{\Omega}(n)=\begin{cases}\Omega+2\tan^{-1}\dfrac{\alpha\sin\Omega}{1-\alpha\cos\Omega} & (0\le n<N_w)\\ \dfrac{n-N_w}{N-N_w}\pi+\dfrac{\pi}{2} & (N_w\le n<N)\end{cases} \quad \cdots (2)$$

ここで、αは伸縮係数、Ωは周波数［ｒａｄ］、N_wはΩ（N_w）＝π／２を満たす値である。

【００１４】

また、ＳＢＭは、上記のような特徴を持つ基底の重み付け線形結合により、ｔフレーム目のスペクトル包絡χｔ＝［χｔ（１），χｔ（１），・・・，χｔ（ｋ），・・・，χｔ（Ｋ）］^{T}を、下記式（３）のように表現する。

【数３】

$$X_t=\exp\left(\frac{1}{2}\phi c_t\right) \quad \cdots (3)$$

ここで、ｃｔ＝［ｃｔ（０），ｃｔ（２），・・・，ｃｔ（ｎ），・・・，ｃｔ（Ｎ－１）］^{T}は、ＳＢＭの基底ベクトルに対するｔフレーム目の重みベクトルであり、Φ＝［Φ０，Φ１，・・・，Φｎ，・・・，ΦＮ－１］は基底ベクトルを行列化したものである。

【００１５】

本実施形態では、ＳＢＭの各基底ベクトルに対応する重みベクトルｃｔを、音声パラメータとして扱う。この音声パラメータは、参考文献１に記載されている非負最小二乗誤差

法を用いて、スペクトル包絡χｔから抽出することができる。すなわち、音声パラメータとしての重みベクトルｃｔは、音声パラメータの値が必ず零以上になるとの制約のもとで、各基底ベクトルと重みベクトルｃｔとの線形結合と、スペクトル包絡χｔと、の誤差が最小となるように最適化を行うことで求められる。

【００１６】

本実施形態では、スペクトル包絡χｔの分析に用いた分析点数が１６０以上であることを想定し、ＳＢＭの基底の数を８０とする。これらの基底のうち、周波数軸上で０ラジアンからπ／２ラジアンまでの低い周波数帯域を表現する１番目の基底から５５番目の基底までは、メルケプストラム分析で用いられるオールパスフィルタの伸縮係数値（ここでは０．３５）に基づいたメル尺度で作成する。また、周波数軸上でπ／２ラジアン以上の高い周波数帯域を表現する５６番目から８０番目の基底は、線形尺度に基づいて作成されたものを用いる。なお、上述した低い周波数帯域の基底は、メル尺度以外の尺度、例えば線形尺度やバーク尺度、ＥＲＢ尺度などを用いて作成されたものを用いてもよい。

【００１７】

なお、本実施形態では、スペクトル包絡χｔから音声パラメータを抽出するための基底モデル１０としてＳＢＭを用いている。しかし、スペクトル包絡χｔから、細分化された局所的な周波数帯域ごとの音声成分を表現した音声パラメータを抽出でき、かつ、抽出した音声パラメータから元のスペクトル包絡χｔを再現できるものであれば、どのような基底モデル１０を用いてもよい。例えば、スパースコーディング法により求めた基底モデルや、非負値行列分解によって求めた基底行列を、スペクトル包絡χｔから音声パラメータを抽出するための基底モデル１０として用いることができる。また、スペクトル包絡χｔから、細分化された局所的な周波数帯域ごとの音声成分を表現した音声パラメータを抽出でき、かつ、抽出した音声パラメータから元のスペクトル包絡χｔを再現できるのであれば、サブバンド分割やフィルタバンクによる表現を用いて、音声パラメータを抽出してもよい。

【００１８】

検出部２は、入力音声のスペクトル包絡χｔ＿ｉｎ、または、このスペクトル包絡χｔ＿ｉｎから抽出部１によって抽出された音声パラメータの包絡形状を解析し、入力音声のスペクトル包絡χｔ＿ｉｎにおいて音声成分が欠損している周波数帯域である欠損帯域を検出する（図２のステップＳ１０２）。

【００１９】

検出部２は、例えば、入力音声のスペクトル包絡χｔ＿ｉｎ、または、このスペクトル包絡χｔ＿ｉｎから抽出された音声パラメータに対して、周波数軸方向の１次の変化の割合および２次の変化の割合を用いて、欠損帯域を検出することができる。

【００２０】

図３は、検出部２による欠損帯域の検出方法の一例を示す図である。図３に示す例は、入力音声が低域通過特性を持つ伝送路を通過することで高周波側の成分が欠損した場合の例であり、スペクトル包絡χｔ＿ｉｎから抽出された音声パラメータの包絡形状を解析して欠損帯域を検出する例である。図の横軸は周波数軸であり、数値は基底の番号を表している。図３（ａ）は、入力音声のスペクトル包絡χｔ＿ｉｎから抽出部１により抽出された音声パラメータの周波数軸方向の変化を表すグラフ図であり、縦軸は音声パラメータの値を示している。また、図３（ｂ）は、図３（ａ）に示した音声パラメータの周波数軸方向の１次変化の割合を表すグラフ図であり、縦軸は音声パラメータを１次微分した値を示している。また、図３（ｂ）は、図３（ａ）に示した音声パラメータの周波数軸方向の２次変化の割合を表すグラフ図であり、縦軸は音声パラメータを２階微分した値を示している。

【００２１】

検出部２は、まず、図３（ｂ）に示す音声パラメータの１次の変化の割合から、値が最小となる次元（以下、第１の基準位置という。）を、次元が大きい方から探索して決定する。次に、検出部２は、第１の基準位置とこの位置から数次元小さい次元との間の範囲を

探索範囲として、図３（ｃ）に示す音声パラメータの２次の変化の割合から、探索範囲内で値が最小となる次元（以下、第２の基準位置という。）を求める。そして、検出部２は、第２の基準点より１つ小さい次元の位置を、欠損帯域の低周波側の端部である開始位置とする。また、図３に示す例では、高周波側の成分が欠損している場合を想定しているため、欠損帯域の高周波側の端部である終了位置は、最大の次元の位置とする。検出部２は、上記のように決定された開始位置と終了位置との間の周波数帯域を、欠損帯域として検出することができる。

【００２２】

入力音声が高域通過特性を持つ伝送路を通過することで低周波側の成分が欠損している場合には、次元の小さい方から上記と同様の処理を行うことで、欠損帯域を検出することができる。すなわち、検出部２は、まず、音声パラメータの１次の変化の割合を次元が小さいほうから探索して、第１の基準位置を決定する。次に、検出部２は、第１の基準位置とこの位置から数次元大きい次元との間の範囲を探索範囲として、音声パラメータの２次の変化の割合から、第２の基準位置を求める。そして、検出部２は、第２の基準位置より１つ大きい次元の位置を、欠損帯域の高周波側の端部である終了位置とする。また、この場合は、欠損帯域の低周波側の端部である開始位置は、最小の次元の位置とする。検出部２は、上記のように決定された開始位置と終了位置との間の周波数帯域を、欠損帯域として検出することができる。

【００２３】

また、入力音声が帯域遮断特性を持つ伝送路を通過することで、低周波と高周波の間の任意の周波数帯域の成分が欠損している場合には、検出部２は、例えば以下の方法で欠損帯域を検出することができる。すなわち、検出部２は、まず、スペクトル傾斜情報を取り除いた音声パラメータに対して、低次元側からの１次の変化の割合および２次の変化の割合を求め、１次の変化の割合の最小値および最大値となる次元をそれぞれ求めて、これらを第１の基準位置とする。次に、検出部２は、最小値となる第１の基準位置より小さい次元において２次の変化の割合が最小となる点を求める。同様に、検出部２は、最大値となる第１の基準位置より大きな次元において変化の割合が最小となる点を求め、それぞれを第２の基準位置とする。そして、検出部２は、これら２つの第２の基準位置に基づいて、低次元側を開始位置、高次元側を終了位置として定める。検出部２は、上記のように定めた開始位置と終了位置との間の周波数帯域を、欠損帯域として検出することができる。

【００２４】

入力音声の伝送路の特性によって欠損帯域が生じる場合、欠損帯域は入力音声ごとに一定であることが想定される。したがって、検出部２は、入力音声の少なくとも１つのフレームに対して上述した処理を行うことで、欠損帯域の検出が可能である。ただし、検出部２は、入力音声の複数のフレームを対象として対して上述した処理を行うようにすれば、欠損帯域の検出をより精度よく行うことができる。この場合、検出部２は、例えば、複数フレームの音声パラメータの平均値を次元ごとに求め、求めた平均値の１次の変化の割合および２次の変化の割合を用いて、欠損位置を精度よく検出することができる。また、検出部２は、複数フレームの音声パラメータに対してそれぞれ上述した処理をそれぞれ行って、得られた結果をマージすることで、最終的な欠損帯域を検出するようにしてもよい。

【００２５】

また、検出部２は、入力音声の各フレームに対して上述した処理を繰り返し行うようにすれば、突発的な要因によって入力音声における欠損帯域がフレーム間で異なる場合であっても、フレーム間で異なる欠損位置をそれぞれ検出することができる。

【００２６】

なお、上述した処理は、入力音声のスペクトル包絡χｔ＿ｉｎから抽出された音声パラメータを処理対象としたが、入力音声のスペクトル包絡χｔ＿ｉｎそのものを処理対象としても、同様の処理によって欠損帯域を検出することができる。すなわち、入力音声のスペクトル包絡χｔ＿ｉｎに対して、周波数軸方向の１次の変化の割合および２次の変化の割合を用いて上記と同様の処理を行うようにしても、欠損帯域を検出することができる。

【００２７】
生成部３は、検出部２により検出された欠損帯域の位置と、統計情報２０と、入力音声のスペクトル包絡χｔ＿ｉｎから抽出部１によって抽出された音声パラメータとに基づいて、欠損帯域に対応する音声パラメータを生成する（図２のステップＳ１０３）。
【００２８】
統計情報２０は、音声成分が欠損していない音声のスペクトル包絡から抽出された音声パラメータ（抽出部１が入力音声のスペクトル包絡χｔ＿ｉｎから抽出する音声パラメータと同様の音声パラメータ）を用いて事前に作成されている。ここで、統計情報とは、音声パラメータベクトルの平均、分散やヒストグラムなどにより、音声パラメータをモデル化したものであり、例えばコードブック、混合分布モデル、隠れマルコフモデルなどである。本実施形態では、統計情報２０として混合正規分布モデル（以下、ＧＭＭという。）を用いる。統計情報２０は、音声処理装置内の図示しない記憶部に予め格納されてもよいし、音声処理装置の動作時に外部から取得されて保持されてもよい。
【００２９】
ＧＭＭでは、重みベクトルｃｔの確率密度関数は、下記式（４）のように表される。
【数４】

$$P(c_t|\lambda)=\sum_{m=1}^{M}P(c_t|m,\lambda)=\sum_{m=1}^{M}\alpha_m N\left(c_t;\mu_m^{(c)},\sum_m^{(cc)}\right) \quad \cdots (4)$$

ここで、λはGMMのパラメータセットを表し、$N\left(c_t;\mu_m^{(c)},\sum_m^{(cc)}\right)$は平均ベクトル$\mu_m^{(c)}$、全共分散行列$\sum_m^{(cc)}$をもつGMMのm番目の正規分布であり、$\alpha_m$はm番目の正規分布の重みである。

【００３０】
なお、本実施形態において、残存帯域（欠損帯域以外の帯域）に対応するパラメータ成分（以下、残存帯域成分という。）の数と、欠損帯域に対応するパラメータ成分（以下、欠損帯域成分という。）の数が異なることを想定している。このため全共分散行列、すなわち、行列のすべての成分にある値を有するものを用いている。しかし、実施形態において残存帯域成分の数と欠損帯域成分の数が常に同数である場合には、全共分散行列の代わりに、行列の対角成分と事前に決定した残存帯域成分とそれに対応する欠損帯域成分とに値を有し、それ以外の成分は零であるような分散行列を用いてもよい。
【００３１】
本実施形態では、音声成分が欠損していない（欠損帯域のない）複数の話者の発話音声から抽出された音声パラメータを学習データとして用いて事前に構築された統計モデルである不特定話者ＧＭＭを、統計情報２０として用いる。統計情報２０の構築には、例えば、ＬＧＢアルゴリズムやＥＭアルゴリズムなどを用いることができる。
【００３２】
生成部３は、統計情報２０としてのＧＭＭを用いて、残存帯域成分から欠損帯域成分を生成するための規則を、次のような手順で求める。
【００３３】
生成部３は、まず、統計情報２０としてのＧＭＭを、検出部２により検出された欠損帯域の位置、すなわち、上述した開始位置および終了位置に基づいて、音声パラメータベクトル、平均ベクトルμｍ（ｃ）、および共分散行列Σｍ（ｃｃ）を分割して、下記式（５）のように変形する。

【数５】

$$P\left(c_t\middle|\lambda\right)=\sum_{m=1}^{M}\alpha_m N\left(\begin{bmatrix}c_t^{(r)}\\ c_t^{(l)}\end{bmatrix};\begin{bmatrix}\mu_m^{(r)}\\ \mu_m^{(l)}\end{bmatrix},\begin{bmatrix}\sum_m^{(rr)}\sum_m^{(rl)}\\ \sum_m^{(lr)}\sum_m^{(ll)}\end{bmatrix}\right) \quad \cdots (5)$$

ここで、$c_t^{(r)}$は残存帯域に関する音声パラメータベクトル、$c_t^{(l)}$は欠損帯域に関する音声パラメータベクトル、$\mu_m^{(r)}$は残存帯域に関する平均ベクトル、$\mu_m^{(l)}$は欠損帯域に関する平均ベクトル、$\sum_m^{(rr)}$は残存帯域に関する自己共分散行列、$\sum_m^{(ll)}$は欠損帯域に関する自己共分散行列、$\sum_m^{(lr)}$は欠損帯域と残存帯域に関する相互共分散行列である。

【００３４】

次に、生成部３は、この変形したＧＭＭを、下記式（６）に示すように、残存帯域の音声パラメータベクトルに対する欠損帯域の音声パラメータベクトルの条件付き確率分布へと変形する。そして、生成部３は、式（６）に示す条件付き確率分布を規則として用いて、残存帯域成分（入力音声のスペクトル包絡χｔ＿ｉｎから抽出された音声パラメータ）から、欠損帯域成分（欠損帯域に対応する音声パラメータ）を生成する。

【数６】

$$P\left(c_t^{(l)}\middle|c_t^{(r)},\lambda\right)=\sum_{m=1}^{M}P\left(m\middle|c_t^{(r)},\lambda\right)P\left(c_t^{(l)}\middle|c_t^{(r)},m,\lambda\right) \quad \cdots (6)$$

ここで、

$$P\left(m\middle|c_t^{(r)},\lambda\right)=\frac{\alpha_m N\left(c_t^{(r)};\mu_m^{(r)},\sum_m^{(rr)}\right)}{\sum_{m=1}^{M}\alpha_m N\left(c_t^{(r)};\mu_m^{(r)},\sum_m^{(rr)}\right)} \quad \cdots (7)$$

$$P\left(c_t^{(l)}\middle|c_t^{(r)},m,\lambda\right)=N\left(c_t^{(l)},E_{m,t}^{(l)},D_m^{(ll)}\right) \quad \cdots (8)$$

$$E_{m,t}^{(l)}=\sum_m^{(lr)}\sum_m^{(rr)^{-1}}\left(c_t^{(r)}-\mu_m^{(r)}\right)+\mu_m^{(l)} \quad \cdots (9)$$

$$D_m^{(l)}=\sum_m^{(ll)}-\sum_m^{(lr)}\sum_m^{(rr)^{-1}}\sum_m^{(rl)} \quad \cdots (10)$$

である。

これより欠損帯域の音声パラメータ$\tilde{c}_t^{(l)}$は最小二乗誤差基準により、下記式(11)のように求まる。

$$\tilde{c}_t^{(l)}=\sum_{m=1}^{M}P\left(m\middle|c_t^{(r)},\lambda\right)E_{m,t}^{(l)} \quad \cdots (11)$$

【００３５】

本実施形態においては、上述したように、１つの入力音声における欠損帯域がフレーム間で一定であることを想定している。この場合、上述したように、フレームごとに欠損帯域に対応する音声パラメータを生成すると、フレーム間で不連続が生じることが考えられる。そこで、この不連続を緩和させるために、生成部３は、当該フレームと前後数フレー

ムを用いて移動平均フィルタ、中央値フィルタ、加重平均フィルタ、ガウスフィルタなどにより平滑化処理を行うことで、欠損帯域に対応する音声パラメータのフレーム間における不連続性を緩和させてもよい。

【００３６】

また、生成部３により生成された欠損帯域に対応する音声パラメータは、汎化されたＧＭＭの影響により平滑化されている。そのため、生成部３は、欠損帯域に対応する音声パラメータを生成した後に、下記の参考文献２で示される系列内変動（以下、ＧＶという）の統計情報や音声パラメータのヒストグラムを用いたパラメータ強調を行ってもよい。

参考文献２：藤敦渉、他４名，「ＧＭＭに基づく最尤変換法による携帯電話音声の帯域拡張」，社団法人　情報処理学会　研究報告（ＩＰＳＪ　ＳＩＧ　Ｔｅｃｈｎｉｃａｌ　Ｒｅｐｏｒｔ），２００７年７月２１日，ｐ．６３－６８

【００３７】

さらに、生成部３は、フレーム間の不連続性や音声パラメータの平滑化を防ぐために、参考文献２で示されている、動的特徴量を用いた尤度最大化基準によるＧＭＭ変換手法を用いて、欠損帯域に対応する音声パラメータを生成してもよい。この場合、ＧＭＭの学習においては、音声パラメータである重みベクトルｃｔと、この重みベクトルｃｔの時間変化成分Δｃｔとを結合した下記式（１２）で示す特徴量Ｃｔを用意し、下記式（１３）に示すＧＭＭを構築して、これを統計情報２０として保持する。

【数７】

$$C_t = \left[c_t^{\Gamma}, \Delta c_t^{\Gamma}\right]^{\Gamma} \quad \cdots (12)$$

$$P(C_t|\lambda) = \sum_{m=1}^{M} P(C_t|m,\lambda) = \sum_{m=1}^{M} \alpha_m N\left(C_t; \mu_m^{(CC)}, \sum_m^{(CC)}\right) \quad \cdots (13)$$

ここで、$\mu_m^{(C)}$はm番目の分布が保持する 結合特徴量の平均ベクトルであり、$\sum_m^{(CC)}$ はm番目の分布が保持する 結合特徴量の全共分散行列である。

【００３８】

式（１３）に示すＧＭＭを統計情報２０として用いる場合においても、生成部３は、まず、検出部２により検出された欠損帯域の位置（開始位置および終了位置）に基づいてＧＭＭを残存帯域成分と欠損帯域成分とに分割し、式（１３）を下記式（１４）のように変形する。

【数８】

$$P(C_t|\lambda) = \sum_{m=1}^{M} \alpha_m N\left(\begin{bmatrix} C_t^{(R)} \\ C_t^{(L)} \end{bmatrix}; \begin{bmatrix} \mu_m^{(R)} \\ \mu_m^{(L)} \end{bmatrix}, \begin{bmatrix} \sum_m^{(RR)} & \sum_m^{(RL)} \\ \sum_m^{(LR)} & \sum_m^{(LL)} \end{bmatrix}\right) \quad \cdots (14)$$

【００３９】

次に、生成部３は、式（１４）に示すＧＭＭを、下記式（１５）に示すように、残存帯域の音声パラメータベクトルに対する欠損帯域の音声パラメータベクトルの条件付き確率分布へと変形する。

【数９】

$$P\left(C_t^{(L)} \middle| C_t^{(R)}, \lambda\right) = \sum_{m=1}^{M} P\left(m \middle| C_t^{(R)}, \lambda\right) P\left(C_t^{(R)} \middle| C_t^{(R)}, m, \lambda\right) \quad \cdots (15)$$

【００４０】

そして、生成部３は、尤度最大化基準で、下記式（１６）および下記式（１７）に示す

ように、欠損帯域の音声パラメータを生成する。

【数１０】

$$\left[\tilde{c}_1^{(l)\Gamma},\tilde{c}_2^{(l)\Gamma},\cdots\tilde{c}_T^{(l)\Gamma}\right]^{\Gamma}=\arg\max_{c^{(l)}}\prod_{t}^{T}\sum_{m=1}^{M}P\left(m\mid C_t^{(R)},\lambda\right)P\left(C_t^{(L)},C_t^{(R)},m,\lambda\right)\quad\cdots(16)$$

$$\text{Subject to}\left[C_1^{(L)\Gamma},C_2^{(L)\Gamma},\cdots C_T^{(L)\Gamma}\right]^{\Gamma}=W\left[c_1^{(l)\Gamma},c_2^{(l)\Gamma},\cdots c_T^{(l)\Gamma}\right]^{\Gamma}\quad\cdots(17)$$

ここで、Ｗは音声パラメータ系列から音声パラメータと時間変化量成分との結合特徴量系列へと変換するための行列を表す。

【００４１】

また、生成部３は、式（１６）の代わりに、参考文献２で示される準最尤分布系列からのパラメータ生成やＧＶを用いたパラメータ生成法を用いて、欠損帯域に対応する音声パラメータを生成してもよいし、式（１６）による音声パラメータの生成後に、ＧＶやヒストグラムを用いたパラメータ強調を行ってもよい。

【００４２】

なお、本実施形態では、統計情報２０として不特定話者ＧＭＭを使用することを想定している。しかし、不特定話者ＧＭＭのほかに、複数の特定話者ＧＭＭを統計情報２０として用いてもよい。この場合、生成部３は、入力音声のスペクトル包絡χｔ＿ｉｎから抽出された音声パラメータに最も適合した特定話者ＧＭＭ、または適合度に合わせて複数の特定話者ＧＭＭを線形結合したものを用いて、欠損帯域に対応する音声パラメータの生成を行う。これにより、欠損帯域の音声パラメータを、入力音声のスペクトル包絡χｔ＿ｉｎから抽出された音声パラメータに適合するように生成することができる。

【００４３】

さらに、入力音声のスペクトル包絡χｔ＿ｉｎから抽出された音声パラメータとの適合性を向上させるために、不特定話者ＧＭＭないしは特定話者ＧＭＭに対して、線形回帰や最大事後確率推定などの統計的な音声認識や音声合成で用いられている話者適応手法を適用し、入力音声のスペクトル包絡χｔ＿ｉｎから抽出された音声パラメータと適合したＧＭＭを用いて、欠損帯域に対応する音声パラメータを生成してもよい。

【００４４】

変換部４は、生成部３が生成した欠損帯域に対応する音声パラメータを、基底モデル１０を用いて、欠損帯域のスペクトル包絡に変換する（図２のステップＳ１０４）。

【００４５】

本実施形態では、基底モデル１０としてＳＢＭを用いるため、上記式（３）に示したような処理を行うことで、欠損帯域に対応する音声パラメータとして生成された重みベクトルｃｔを、欠損帯域の音声スペクトル包絡χ~ｔに変換することができる。すなわち、変換部４は、欠損帯域に対応する音声パラメータである重みベクトルｃｔと、この欠損帯域に対応する基底ベクトルとを線形結合することにより、欠損帯域のスペクトル包絡χ~ｔを求めることができる。

【００４６】

補完部５は、変換部４により得られた欠損帯域のスペクトル包絡χ~ｔと、入力音声のスペクトル包絡χｔ＿ｉｎとを合成して、欠損帯域が補完されたスペクトル包絡χｔ＿ｏｕｔを生成する（図２のステップＳ１０５）。

【００４７】

補完部５は、例えば、入力音声のスペクトル包絡χｔ＿ｉｎのうち、検出部２により検出された欠損帯域の位置（開始位置と終了位置との間の帯域）に、変換部４により得られた欠損帯域のスペクトル包絡χ~ｔを当てはめるとともに、不連続性を緩和させる処理を行ってこれらを合成することで、欠損帯域が補完されたスペクトル包絡χｔ＿ｏｕｔを生成することができる。

【００４８】
図４は、補完部５による処理の一例を示す図である。図４に示す例は、低域通過特性を持つ伝送路により高周波側の成分が欠損した入力音声のスペクトル包絡χｔ＿ｉｎから、欠損帯域が補完されたスペクトル包絡χｔ＿ｏｕｔを生成する例である。
【００４９】
入力音声のスペクトル包絡χｔ＿ｉｎの欠損帯域の位置に、変換部４により得られた欠損帯域のスペクトル包絡χ~ｔをそのまま当てはめると、欠損帯域の境界位置にて２つのスペクトル包絡の値が大きくずれて、不連続性が発生する場合がある。そこで、補完部５は、まず、欠損帯域の境界位置における２つのスペクトル包絡の差分ｄを計測する（図４（ａ））。そして、補完部５は、計測した差分ｄに基づき、変換部４により得られた欠損帯域のスペクトル包絡χ~ｔの全体にバイアス補正を行う（図４（ｂ））。
【００５０】
次に、補完部５は、入力音声のスペクトル包絡χｔ＿ｉｎと欠損帯域のスペクトル包絡χ~ｔとが滑らかに接続されるように、それぞれのスペクトル包絡の境界位置周辺の成分に対して片側ハン窓をかけ（図４（ｃ））、該当する箇所のスペクトル包絡の成分を加算することで、入力音声のスペクトル包絡χｔ＿ｉｎと欠損帯域のスペクトル包絡χ~ｔとを合成する（図４（ｄ））。これにより、欠損帯域が補完されたスペクトル包絡χｔ＿ｏｕｔが生成される。
【００５１】
なお、高域通過特性を持つ伝送路により低周波側の成分が欠損した入力音声のスペクトル包絡χｔ＿ｉｎから、欠損帯域が補完されたスペクトル包絡χｔ＿ｏｕｔを生成する場合も、上記と同様の手順で、欠損帯域が補完されたスペクトル包絡χｔ＿ｏｕｔを適切に生成することができる。
【００５２】
図５は、補完部５による処理の他の例を示す図である。図５に示す例は、帯域遮断特性を持つ伝送路により低周波と高周波の間の任意の周波数帯域の成分が欠損した入力音声のスペクトル包絡χｔ＿ｉｎから、欠損帯域が補完されたスペクトル包絡χｔ＿ｏｕｔを生成する例である。
【００５３】
図５の例の場合、補完部５は、欠損帯域の開始位置における２つのスペクトル包絡の差分ｄｓを計測するとともに、欠損帯域の終了位置における２つのスペクトル包絡の差分ｄｅを計測する（図５（ａ））。そして、補完部５は、欠損帯域の開始位置で計測された差分ｄｓと、欠損帯域の終了位置で計測された差分ｄｅとに基づき、変換部４により得られた欠損帯域のスペクトル包絡χ~ｔに対して傾斜補正をかける（図５（ｂ））。
【００５４】
次に、補完部５は、欠損帯域の開始位置と終了位置の双方において、入力音声のスペクトル包絡χｔ＿ｉｎと欠損帯域のスペクトル包絡χ~ｔとが滑らかに接続されるように、これら開始位置および終了位置の周辺におけるそれぞれのスペクトル包絡の成分に対して片側ハン窓をかけ（図５（ｃ））、該当する箇所のスペクトル包絡の成分を加算することで、入力音声のスペクトル包絡χｔ＿ｉｎと欠損帯域のスペクトル包絡χ~ｔとを合成する（図５（ｄ））。これにより、欠損帯域が補完されたスペクトル包絡χｔ＿ｏｕｔが生成される。
【００５５】
本実施形態の音声処理装置は、補完部５により生成された、欠損帯域が補完されたスペクトル包絡χｔ＿ｏｕｔを外部に出力することができる。また、本実施形態の音声処理装置は、欠損帯域が補完されたスペクトル包絡χｔ＿ｏｕｔから音声を復元し、復元した音声を出力するようにしてもよい。
【００５６】
以上、具体的な例を挙げながら詳細に説明したように、本実施形態の音声処理装置によれば、任意の周波数帯域で欠損した音声成分を適切に補完することができる。

【００５７】
なお、本実施形態の音声処理装置は、例えば、汎用のコンピュータ装置を基本ハードウェアとして用いて実現することが可能である。すなわち、本実施形態の音声処理装置は、汎用のコンピュータ装置に搭載されたプロセッサにプログラムを実行させることにより実現することができる。このとき、音声処理装置は、上記のプログラムをコンピュータ装置にあらかじめインストールすることで実現してもよいし、ＣＤ－ＲＯＭなどの記憶媒体に記憶して、あるいはネットワークを介して上記のプログラムを配布して、このプログラムをコンピュータ装置に適宜インストールすることで実現してもよい。また、上記のプログラムをサーバーコンピュータ装置上で実行させ、ネットワークを介してその結果をクライアントコンピュータ装置で受け取ることにより実現してもよい。

【００５８】
また、本実施形態の音声処理装置で使用する各種情報は、上記のコンピュータ装置に内蔵あるいは外付けされたメモリ、ハードディスクもしくはＣＤ－Ｒ、ＣＤ－ＲＷ、ＤＶＤ－ＲＡＭ、ＤＶＤ－Ｒなどの記録媒体を適宜利用して格納しておくことができる。例えば、本実施形態の音声処理装置が使用する基底モデル１０や統計情報２０は、これら記録媒体を適宜利用して格納しておくことができる。

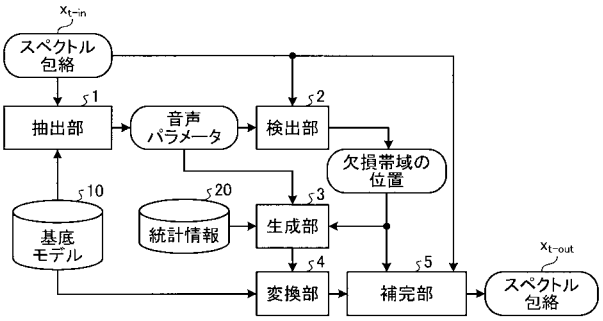
【００５９】
本実施形態の音声処理装置で実行されるプログラムは、音声処理装置を構成する各処理部（抽出部１、検出部２、生成部３、変換部４および補完部５）を含むモジュール構成となっており、実際のハードウェアとしては、例えば、プロセッサが上記記憶媒体からプログラムを読み出して実行することにより、上記各処理部が主記憶装置上にロードされ、主記憶装置上に生成されるようになっている。

【００６０】
以上、本発明の実施形態を説明したが、ここで説明した実施形態は、例として提示したものであり、発明の範囲を限定することは意図していない。ここで説明した新規な実施形態は、その他の様々な形態で実施されることが可能であり、発明の要旨を逸脱しない範囲で、種々の省略、置き換え、変更を行うことができる。ここで説明した実施形態やその変形は、発明の範囲や要旨に含まれるとともに、特許請求の範囲に記載された発明とその均等の範囲に含まれる。

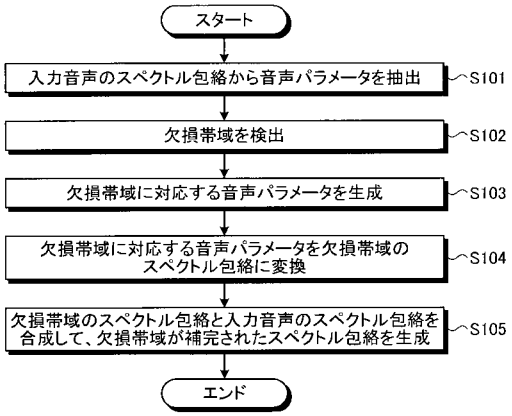
【符号の説明】

【００６１】
１　抽出部
２　検出部
３　生成部
４　変換部
５　補完部
１０　基底モデル
２０　統計情報

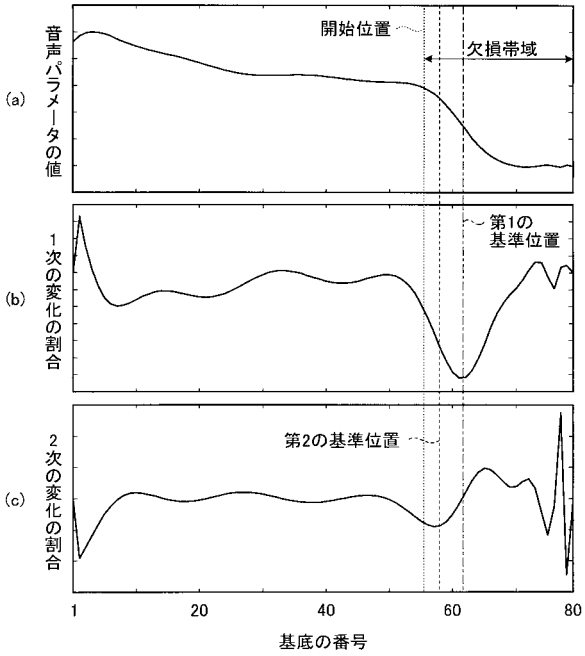
【図１】


x_{t-in}
スペクトル包絡
抽出部
音声パラメータ
検出部
欠損帯域の位置
基底モデル
統計情報
生成部
変換部
補完部
x_{t-out}
スペクトル包絡
1
2
3
4
5
10
20

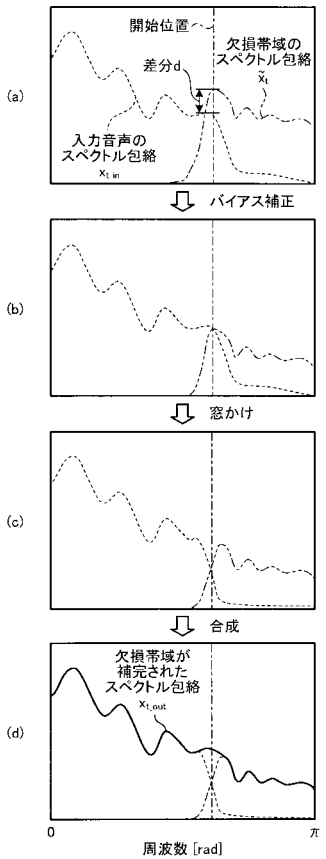
【図２】


スタート
入力音声のスペクトル包絡から音声パラメータを抽出
S101
欠損帯域を検出
S102
欠損帯域に対応する音声パラメータを生成
S103
欠損帯域に対応する音声パラメータを欠損帯域のスペクトル包絡に変換
S104
欠損帯域のスペクトル包絡と入力音声のスペクトル包絡を合成して、欠損帯域が補完されたスペクトル包絡を生成
S105
エンド

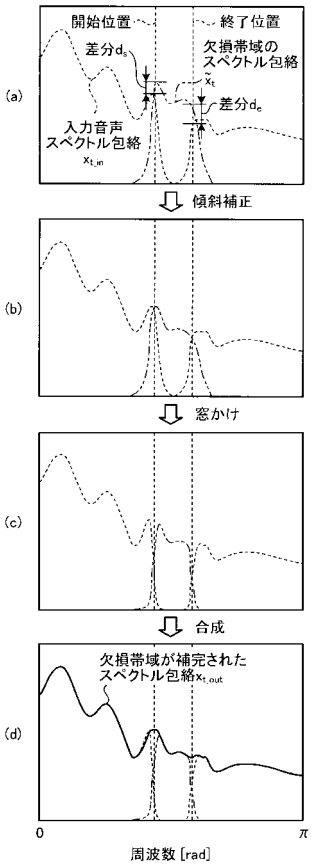
【図３】


(a)
音声パラメータの値
開始位置
欠損帯域
(b)
1次の変化の割合
第1の基準位置
(c)
2次の変化の割合
第2の基準位置
1
20
40
60
80
基底の番号

【図４】


(a)
開始位置
差分d
欠損帯域のスペクトル包絡
$\tilde{x}_t$
入力音声のスペクトル包絡
x_{t_in}
バイアス補正
(b)
窓かけ
(c)
合成
(d)
欠損帯域が補完されたスペクトル包絡
x_{t_out}
0
π
周波数 [rad]

【図５】


(a)
開始位置
終了位置
差分d_s
欠損帯域のスペクトル包絡
$\tilde{x}_t$
差分d_e
入力音声スペクトル包絡
x_{t_in}
傾斜補正
(b)
窓かけ
(c)
合成
(d)
欠損帯域が補完されたスペクトル包絡x_{t_out}
0
π
周波数 [rad]