(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第5229234号
(P5229234)
(45)発行日　平成25年7月3日(2013.7.3)　(24)登録日　平成25年3月29日(2013.3.29)

(51)Int.Cl.　F I
G10L 15/04 (2013.01)　G10L 15/04 300B
G10L 25/78 (2013.01)　G10L 11/02

請求項の数 7　(全 32 頁)

(21)出願番号　特願2009-546107(P2009-546107)
(86)(22)出願日　平成19年12月18日(2007.12.18)
(86)国際出願番号　PCT/JP2007/074274
(87)国際公開番号　WO2009/078093
(87)国際公開日　平成21年6月25日(2009.6.25)
審査請求日　平成22年2月16日(2010.2.16)

前置審査

(73)特許権者　000005223
富士通株式会社
神奈川県川崎市中原区上小田中４丁目１番１号
(74)代理人　100078868
弁理士　河野　登夫
(72)発明者　鷲尾　信之
神奈川県川崎市中原区上小田中４丁目１番１号　富士通株式会社内
(72)発明者　早川　昭二
神奈川県川崎市中原区上小田中４丁目１番１号　富士通株式会社内

審査官　菊池　智紀

最終頁に続く

(54)【発明の名称】非音声区間検出方法及び非音声区間検出装置

(57)【特許請求の範囲】
【請求項１】
音を標本化した音データから所定の時間長の複数のフレームを生成し、人が発声した音声に基づく音声データを含まないフレームを有する非音声区間を検出する非音声区間検出方法において、
各フレームの音データを周波数軸上の成分に変換したスペクトルについて、０次の自己相関関数に対する１次の自己相関関数の比の絶対値を導出し、
導出した絶対値が、所定の閾値以上であるか否かを判定し、
前記閾値以上であると判定したフレームが連なる数を計数し、
計数した数が前記閾値に応じて定める所定数以上であるか否かを判定し、
所定数以上であると判定したときに、前記フレームが連なる区間を非音声区間として検出する
ことを特徴とする非音声区間検出方法。
【請求項２】
音を標本化した音データから所定の時間長の複数のフレームを生成し、人が発声した音声に基づく音声データを含まないフレームを有する非音声区間を検出する非音声区間検出方法において、
各フレームの音データを周波数軸上の成分に変換したスペクトルについて、０次の自己相関関数に対する１次の自己相関の比を導出し、
導出した比について、前フレームとの変化量の絶対値を導出し、

導出した変化量の絶対値が、所定の閾値以下であるか否かを判定し、
前記閾値以下であると判定したフレームが連なる数を計数し、
計数した数が前記閾値に応じて定める所定数以上であるか否かを判定し、
所定数以上であると判定したときに、前記フレームが連なる区間を非音声区間として検出する
ことを特徴とする非音声区間検出方法。

【請求項３】

音を標本化した音データから所定の時間長の複数のフレームを生成し、人が発声した音声に基づく音声データを含まないフレームを有する非音声区間を検出する非音声区間検出装置において、
各フレームの音データを周波数軸上の成分に変換したスペクトルについて、０次の自己相関関数に対する１次の自己相関関数の比の絶対値を導出する導出手段と、
導出した絶対値が、所定の閾値以上であるか否かを判定する判定手段と、
前記閾値以上であると判定したフレームが連なる数を計数する手段と、
計数した数が前記閾値に応じて定める所定数以上であるか否かを判定する手段と、
所定数以上であると判定したときに、前記フレームが連なる区間を非音声区間として検出する検出手段と
を備えることを特徴とする非音声区間検出装置。

【請求項４】

音を標本化した音データから所定の時間長の複数のフレームを生成し、人が発声した音声に基づく音声データを含まないフレームを有する非音声区間を検出する非音声区間検出装置において、
各フレームの音データを周波数軸上の成分に変換したスペクトルについて、０次の自己相関関数に対する１次の自己相関の比を導出する導出手段と、
導出した比について、前フレームとの変化量の絶対値を導出する第２の導出手段と、
導出した変化量の絶対値が所定の閾値以下であるか否かを判定する判定手段と、
前記閾値以下であると判定したフレームが連なる数を計数する手段と、
計数した数が前記閾値に応じて定める所定数以上であるか否かを判定する手段と、
所定数以上であると判定したときに、前記フレームが連なる区間を非音声区間として検出する検出手段と
を備えることを特徴とする非音声区間検出装置。

【請求項５】

前記第２の導出手段が導出した変化量が、前記閾値より大きい第２の閾値を超えるか否かを判定する第２の判定手段を備え、
前記検出手段は、前記第２の判定手段が第２の閾値を超えると判定した場合、該判定が成立するフレームを含めて第２の所定数だけ連なるフレームからなる区間を、非音声区間の検出対象から除外するように構成してあることを特徴とする請求項４に記載の非音声区間検出装置。

【請求項６】

前記第２の判定手段の判定が成立するフレームが連なる数を計数する手段と、
計数した数が所定数以下であるか否かを判定する手段と、
所定数以下であると判定した場合、該判定が成立するフレーム及び前記第２の所定数未満のフレームが連なる区間が、非音声区間に挟まれているときに、前記非音声区間に挟まれた区間を非音声区間として検出する第２の検出手段と
を備えることを特徴とする請求項５に記載の非音声区間検出装置。

【請求項７】

前記尺度は、音データのＮ次（Ｎは０以上の整数）の自己相関関数に対するＭ次（ＭはＮと異なる０以上の整数）の自己相関関数の比であることを特徴とする請求項３乃至６の何れかに記載の非音声区間検出装置。

【発明の詳細な説明】

【技術分野】
【０００１】
本発明は、音を標本化した音データから所定の時間長のフレームを生成し、非音声区間を検出する非音声区間検出方法、該非音声区間検出方法を適用した非音声区間検出装置に関し、特に非音声の特徴を有する物理量と所定の閾値との比較に基づいて、非音声区間を検出する非音声区間検出方法及び非音声区間検出装置に関する。
【背景技術】
【０００２】
カーナビゲーション装置に代表される車載装置に多く用いられる音声認識装置では、一般的には音声区間を検出し、検出した音声区間について算出した音声の特徴量に基づいて、単語列を認識する。特に音声区間の検出を誤った場合、当該区間における音声の認識率が低下するため、音声区間を的確に検出すること、又は非音声区間を検出して音声認識の対象から除外することが重要である。
【０００３】
音声区間の基本的な検出方式として、入力音声のパワーが、その時の推定背景雑音レベルに閾値を加えた基準値を超えた区間を、音声区間として扱うものがある。この場合は、ブザー音のようにパワー変動が大きい雑音、ワイパーの摺動音、及び音声プロンプトのエコー等、何れも非定常性が強い雑音を含む区間を、音声区間として誤検出する可能性が高い。そこで、直近の発声中の最大音声パワー及びその時の音声認識結果より補正係数を導出し、推定背景雑音レベルと併せて、以後の基準値を補正する技術が、特許文献１に開示されている。
【特許文献１】特開平７－９２９８９号公報
【発明の開示】
【発明が解決しようとする課題】
【０００４】
しかしながら、特許文献１に開示されている技術では、発声前後の非音声区間は除外できても、発声がない場合に基準値を補正することができず、雑音のみの区間を音声区間として誤検出することがある問題は解消されない。
【０００５】
本発明は斯かる事情に鑑みてなされたものであり、音データの周波数スペクトルに偏りを有するフレームが、音声らしからぬ程度に連なる区間、又は周波数スペクトルの偏り、パワー若しくはピッチについての変化に乏しい音データを有するフレームが音声らしからぬ程度に連なる区間を、非音声区間として検出することにより、パワーの大きい雑音若しくは非定常性の強い雑音、又はパワー変動の大きい雑音が発生する環境下においても、発声前後か否かに拘わらず、高精度に非音声区間を検出することが可能な非音声区間検出方法、及び該非音声区間検出方法を適用した非音声区間検出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】
【０００６】
第１の非音声区間検出方法は、音を標本化した音データから所定の時間長の複数のフレームを生成し、人が発声した音声に基づく音声データを含まないフレームを有する非音声区間を検出する非音声区間検出方法において、各フレームの音データを周波数軸上の成分に変換したスペクトルについて、０次の自己相関関数に対する１次の自己相関関数の比の絶対値を導出し、導出した絶対値が、所定の閾値以上であるか否かを判定し、前記閾値以上であると判定したフレームが連なる数を計数し、計数した数が前記閾値に応じて定める所定数以上であるか否かを判定し、所定数以上であると判定したときに、前記フレームが連なる区間を非音声区間として検出することを要件とする。
【０００７】
第２の非音声区間検出方法は、音を標本化した音データから所定の時間長の複数のフレームを生成し、人が発声した音声に基づく音声データを含まないフレームを有する非音声

区間を検出する非音声区間検出方法において、各フレームの音データを周波数軸上の成分に変換したスペクトルについて、０次の自己相関関数に対する１次の自己相関の比を導出し、導出した比について、前フレームとの変化量の絶対値を導出し、導出した変化量の絶対値が、所定の閾値以下であるか否かを判定し、前記閾値以下であると判定したフレームが連なる数を計数し、計数した数が前記閾値に応じて定める所定数以上であるか否かを判定し、所定数以上であると判定したときに、前記フレームが連なる区間を非音声区間として検出することを要件とする。

【０００８】

第３の非音声区間検出装置は、音を標本化した音データから所定の時間長の複数のフレームを生成し、人が発声した音声に基づく音声データを含まないフレームを有する非音声区間を検出する非音声区間検出装置において、各フレームの音データを周波数軸上の成分に変換したスペクトルについて、０次の自己相関関数に対する１次の自己相関関数の比の絶対値を導出する導出手段と、導出した絶対値が、所定の閾値以上であるか否かを判定する判定手段と、前記閾値以上であると判定したフレームが連なる数を計数する手段と、計数した数が前記閾値に応じて定める所定数以上であるか否かを判定する手段と、所定数以上であると判定したときに、前記フレームが連なる区間を非音声区間として検出する検出手段とを備えることを要件とする。

【０００９】

第４の非音声区間検出装置は、音を標本化した音データから所定の時間長の複数のフレームを生成し、人が発声した音声に基づく音声データを含まないフレームを有する非音声区間を検出する非音声区間検出装置において、各フレームの音データを周波数軸上の成分に変換したスペクトルについて、０次の自己相関関数に対する１次の自己相関の比を導出する導出手段と、導出した比について、前フレームとの変化量の絶対値を導出する第２の導出手段と、導出した変化量の絶対値が所定の閾値以下であるか否かを判定する判定手段と、前記閾値以下であると判定したフレームが連なる数を計数する手段と、計数した数が前記閾値に応じて定める所定数以上であるか否かを判定する手段と、所定数以上であると判定したときに、前記フレームが連なる区間を非音声区間として検出する検出手段とを備えることを要件とする。

【００１０】

第５の非音声区間検出装置は、第４の装置において、前記第２の導出手段が導出した変化量が、前記閾値より大きい第２の閾値を超えるか否かを判定する第２の判定手段を備え、前記検出手段は、前記第２の判定手段が第２の閾値を超えると判定した場合、該判定が成立するフレームを含めて第２の所定数だけ連なるフレームからなる区間を、非音声区間の検出対象から除外するように構成してあることを要件とする。

【００１１】

第６の非音声区間検出装置は、第５の装置において、前記第２の判定手段の判定が成立するフレームが連なる数を計数する手段と、計数した数が所定数以下であるか否かを判定する手段と、所定数以下であると判定した場合、該判定が成立するフレーム及び前記第２の所定数未満のフレームが連なる区間が、非音声区間に挟まれているときに、前記非音声区間に挟まれた区間を非音声区間として検出する第２の検出手段とを備えることを要件とする。

【００１２】

本願の非音声区間検出装置は、前記第２の導出手段による変化量の導出の対象となったフレームを含めて、所定数だけ連なるフレームについて、変化量の最大値を導出する第３の導出手段を備え、前記判定手段は、前記第３の導出手段が導出した最大値を、前記第２の導出手段が導出した変化量として扱うように構成してあることを要件とする。

【００１３】

第７の非音声区間検出装置は、第３の装置乃至第６の装置の何れかにおいて、前記尺度は、音データのＮ次（Ｎは０以上の整数）の自己相関関数に対するＭ次（ＭはＮと異なる０以上の整数）の自己相関関数の比であることを要件とする。

【００１４】
本願の非音声区間検出装置は、前記導出手段が、各フレームについてスペクトルの偏倚を導出した場合、前記各フレームに夫々時系列に前後する複数のフレームについて、スペクトルの偏倚の最大値、最小値、平均値及び中央値の少なくとも一を導出して、導出した値を前記各フレーム夫々についてのスペクトルの偏倚として扱うように構成してあることを要件とする。
【００１５】
本願の非音声区間検出装置は、前記判定手段が判定の対象とした全フレームの数に対する、前記判定が成立するフレームの数の割合を算出する手段と、算出した割合が、所定の割合以上であるか否かを判定する手段と、該判定が成立するフレームが連なる数を計数する手段と、計数した数が所定数以上であるか否かを判定する手段と、所定数以上であると判定したときに、前記フレームが連なる区間を非音声区間として検出する第３の検出手段とを備えることを要件とする。
【００１６】
本願の非音声区間検出装置は、非音声区間として検出されたフレームの音データ、及び前記非音声区間以外のフレームの音データに基づいて、信号対雑音比を導出する手段と、導出した信号対雑音比に基づいて、前記閾値を変更する手段とを備えることを要件とする。
【００１７】
本願の非音声区間検出装置は、各フレームの音データについて、ピッチの各周波数成分の強度の最大値を導出する手段と、導出した強度の最大値に基づいて、前記閾値を変更する手段とを備えることを要件とする。
【００１８】
本願の非音声区間検出装置は、人が発声した音データについて、予め準備された複数の候補閾値に対し、前記判定手段の判定が成立するフレームが連なる個数を夫々集計する手段と、集計した結果に基づいて、複数の候補閾値の中から前記閾値を決定する手段とを備えることを要件とする。
【００１９】
本願の非音声区間検出装置は、各フレームの音データのパワーを導出する第４の導出手段と、各フレームの１又は複数の前フレームの音データのパワーに基づいて、夫々のフレームの背景雑音パワーを推定する推定手段と、各フレームについて前記第４の導出手段が導出したパワーが、夫々のフレームについて前記推定手段が推定した背景雑音パワーより、所定の閾値以上大きいか否かを判定する手段と、前記背景雑音パワーより前記閾値以上大きいと判定したフレームからなる区間を音声区間として検出する第４の検出手段とを備え、前記推定手段は、前記第４の検出手段が検出した音声区間のフレームについて、前フレームの背景雑音パワーを維持するように構成してあり、更に、前記第４の検出手段が検出した音声区間のうち、前記検出手段によって非音声区間として検出されたフレームについて、背景雑音パワーを推定するように構成してあることを要件とする。
【００２０】
本願の非音声区間検出装置は、各フレームの音データのパワーを導出する第４の導出手段と、各フレームの１又は複数の前フレームの音データのパワーに基づいて、夫々のフレームの背景雑音パワーを推定する推定手段と、各フレームについて前記第４の導出手段が導出したパワーが、夫々のフレームについて前記推定手段が推定した背景雑音パワーより、所定の閾値以上大きいか否かを判定する手段と、前記背景雑音パワーより前記閾値以上大きいと判定したフレームからなる区間を音声区間として検出する第４の検出手段とを備え、前記推定手段は、前記第４の検出手段が検出した音声区間のフレームについて、前フレームの背景雑音パワーを維持するように構成してあり、更に、前記第４の検出手段が検出した音声区間の全部又は一部が、前記検出手段によって非音声区間として検出された回数を計数する手段と、計数した回数が所定回数以上であるか否かを判定する手段と、所定回数以上であると判定した場合、該判定が成立した際のフレームの音データのパワーを、

背景雑音パワーとして更新する手段とを備えることを要件とする。

【００２１】

第１の方法及び第３の装置では、音データを周波数軸上の成分に変換したスペクトルにおける高周波側又は低周波側への偏りの大きさを示す尺度が所定の閾値以上となるフレームが所定数以上連なる区間を、非音声区間として検出することにより、音データの周波数スペクトルに偏りを有するフレームが音声らしからぬ程度に連なる区間を非音声区間として検出するので、パワーの大きい雑音又は非定常性の強い雑音が発生する環境下においても、高精度に非音声区間を検出することが可能である。

【００２２】

第２の方法及び第４の装置では、音データの周波数スペクトルにおける高周波側又は低周波側への偏りの大きさを示す尺度、パワー及びピッチの少なくとも一について前フレームとの変化量が所定の閾値以下となるフレームが、所定数以上連なる区間を非音声区間として検出することにより、周波数スペクトルにおける前記尺度、パワー若しくはピッチについての変化に乏しい音データを有するフレームが音声らしからぬ程度に連なる区間を非音声区間として検出するので、パワー変動の大きい雑音が発生する環境下においても、高精度に非音声区間を検出することが可能である。

【００２３】

第５の装置では、導出した指標の前フレームとの変化量が前記閾値より大きい第２の閾値を超えるフレームを含めて第２の所定数だけ連なるフレームからなる区間を、非音声区間として検出することがないので、音声データを含む可能性のあるフレームからなる区間を、非音声区間として誤検出することを防止することが可能である。

【００２４】

第６の装置では、導出した指標の前フレームとの変化量が第２の閾値を超えて所定数以下だけ連なるフレーム及び第２の所定数以下のフレームからなる区間が、非音声区間に挟まれている場合に、その挟まれた区間を非音声区間として検出することにより、音データの単発的な変化が発生した場合であっても、高精度に非音声区間を検出することが可能である。

【００２５】

本願の装置では、連なる所定数のフレームについて、夫々導出した指標の前フレームとの変化量の最大値を、一のフレームについての前フレームとの変化量として扱うことにより、各フレームの指標について当初導出した前フレームとの変化量が近傍のフレームについての当該変化量の最大値と置き換わるので、音声データを含む可能性のあるフレームからなる区間を、非音声区間として誤検出することを抑止することが可能である。

【００２６】

第７の装置では、音データの自己相関関数のＮ次の値に対するＭ次の値の比が、音データのスペクトルの包絡を近似する指標であるので、これをスペクトルにおける高周波側又は低周波側への偏りの大きさを示す尺度とすることにより、音データの周波数スペクトルの偏りが的確に把握されて、高精度に非音声区間を検出することが可能である。

【００２７】

本願の装置では、前後する所定数のフレームについて、夫々導出したスペクトルの偏倚の最大値、最小値、平均値及び中央値の少なくとも一を、一のフレームについてのスペクトルの偏倚として扱うことにより、スペクトルの偏倚が短時間に変化した場合であっても、高精度に非音声区間を検出することが可能である。

【００２８】

本願の装置では、音データの周波数スペクトルの偏倚が正の値（又は負の値）の場合、所定の閾値以上（又は所定の閾値以下）となるフレーム、又は導出した指標の前フレームとの変化量が前記閾値と異なる他の閾値以下となるフレームが、所定の割合以上で所定数以上連なる区間を、非音声区間として検出することにより、音データの周波数スペクトルの偏倚、又は導出した指標の前フレームとの変化量が、短時間に変動する場合にも、高精度に非音声区間を検出することが可能である。

【００２９】
本願の装置では、検出した非音声区間の音データ及び非音声区間以外の音データより導出した信号対雑音比に基づいて、前記閾値を変更することにより、例えば信号対雑音比が低下して、スペクトルの偏倚又は導出した指標の前フレームとの変化量が変動した場合に、前記閾値を適切に調整して、非音声区間の誤検出を抑止することができ、高精度に非音声区間を検出することが可能である。
【００３０】
本願の装置では、ピッチの各周波数成分の強度についての最大値に基づいて、前記閾値を調整することにより、ピッチが明瞭に現れる度合いに応じて前記閾値を適切に調整することができるので、高精度に非音声区間を検出することが可能である。
【００３１】
本願の装置では、予め準備した複数の候補閾値を所定の音声データに適用し、夫々の閾値以上（又は閾値以下）となるフレームが連なる個数を集計した結果に基づいて、前記閾値を決定することにより、事前の学習に基づいて前記閾値を決定することができるので、高精度に非音声区間を検出することが可能である。
【００３２】
本願の装置では、非音声区間のフレームの音データのパワーに基づいて推定した背景雑音パワーより、所定の閾値以上大きいパワーを有するフレームからなる区間を音声区間として検出し、検出した音声区間のうち、非音声区間として検出されたフレームについて、背景雑音パワーを推定するので、音データのパワーに基づいて音声検出した結果を適正に修正することが可能である。
【００３３】
本願の装置では、非音声区間のフレームの音データのパワーに基づいて推定した背景雑音パワーより、所定の閾値以上大きいパワーを有するフレームからなる区間を音声区間として検出し、検出した音声区間の全部又は一部が、所定回数だけ非音声区間として検出された際のフレームの音データのパワーを、背景雑音パワーとして更新するので、背景雑音パワーの推定値が上がり過ぎて、音声区間が検出できなくなることを抑止することができる。
【発明の効果】
【００３４】
開示の非音声区間検出方法、及び非音声区間検出装置は、各フレームの音データを周波数軸上の成分に変換したスペクトルにおける高周波側又は低周波側への偏りの大きさを示す尺度が所定の閾値以上であるかを判定し、前記閾値以上と判定したフレームが連なる数が所定数以上かを判定し、そして所定数以上と判定したフレームが連なる区間を非音声区間として検出する。
【００３５】
この構成により、開示の方法及び装置では、スペクトルの偏りに係る閾値とフレームが連なる数に係る閾値とを組み合わせて、非音声の特徴を有するフレームが音声らしからぬ程度に連なる区間を非音声区間として検出し、人の発声による基準値の補正を要しない。従って、パワーの大きい雑音、又は非定常性の強い雑音が発生する環境下においても、発声前後か否かに拘わらず、高精度に非音声区間を検出することが可能である等、優れた効果を奏する。
【００３６】
また、開示の非音声区間検出方法、及び非音声区間検出装置は、各フレームの音データを周波数軸上の成分に変換したスペクトルのにおける高周波側又は低周波側への偏りの大きさを示す尺度を少なくとも用いて、前フレームとの変化量が所定の閾値以下であるかを判定し、前記閾値以下と判定したフレームが連なる数が所定数以上かを判定し、そして所定数以上と判定したフレームが連なる区間を非音声区間として検出する。
【００３７】
この構成により、開示の方法及び装置では、周波数スペクトルの偏り、パワー若しくは

ピッチについての変化に係る閾値とフレームが連なる数に係る閾値とを組み合わせて、非音声の特徴を有するフレームが音声らしからぬ程度に連なる区間を非音声区間として検出し、人の発声による基準値の補正を要しない。従って、パワー変動の大きい雑音が発生する環境下においても、発声前後か否かに拘わらず、高精度に非音声区間を検出することが可能である等、優れた効果を奏する。

【図面の簡単な説明】

【００３８】

【図１】本発明の実施の形態１に係る非音声区間検出装置の一実施例である音声認識装置の構成例を示すブロック図である。

【図２】制御手段の音声認識に係る処理構成例を示すブロック図である。

【図３】制御手段の音声認識処理の一例を示すフローチャートである。

【図４】非音声区間検出のサブルーチンに係る制御手段の処理手順を示すフローチャートである。

【図５】鼻をすする音について、パワー及び高域・低域強度等のデータを示す図である。

【図６】踏切の警報音について、パワー及び高域・低域強度等のデータを示す図である。

【図７】発声音（「えーテスト中です」）について、パワー及び高域・低域強度等のデータを示す図である。

【図８】発声音（「経営（けーえー）」）について、パワー及び高域・低域強度等のデータを示す図である。

【図９】本発明の実施の形態２に係る非音声区間検出装置の一実施例である音声認識装置について、制御手段の音声認識に係る処理構成例を示すブロック図である。

【図１０】本発明の実施の形態３に係る非音声区間検出装置の一実施例である音声認識装置について、制御手段の音声認識に係る処理構成例を示すブロック図である。

【図１１】制御手段の音声認識処理の一例を示すフローチャートである。

【図１２】非音声区間検出のサブルーチンに係る制御手段の処理手順を示すフローチャートである。

【図１３】非音声区間検出除外のサブルーチンに係る制御手段の処理手順を示すフローチャートである。

【図１４】非音声区間検出除外のサブルーチンに係る制御手段の処理手順を示すフローチャートである。

【図１５】非音声区間検出確定のサブルーチンに係る制御手段の処理手順を示すフローチャートである。

【図１６】非音声区間検出確定のサブルーチンに係る制御手段の処理手順を示すフローチャートである。

【図１７】本発明の実施の形態４に係る非音声検出装置の一実施例である音声認識装置について、非音声区間検出のサブルーチンに係る制御手段の処理手順を示すフローチャートである。

【図１８】本発明の実施の形態４に係る非音声検出装置の一実施例である音声認識装置について、非音声区間検出のサブルーチンに係る制御手段の処理手順を示すフローチャートである。

【図１９】本発明の実施の形態５に係る非音声検出装置の一実施例である音声認識装置について、制御手段の音声認識処理の一例を示すフローチャートである。

【図２０】本発明の実施の形態６に係る非音声検出装置の一実施例である音声認識装置について、非音声区間検出のサブルーチンに係る制御手段の処理手順を示すフローチャートである。

【図２１】本発明の実施の形態６に係る非音声検出装置の一実施例である音声認識装置について、非音声区間検出のサブルーチンに係る制御手段の処理手順を示すフローチャートである。

【図２２】本発明の実施の形態７に係る非音声検出装置の一実施例である音声認識装置について、制御手段の音声認識処理の一例を示すフローチャートである。

【符号の説明】
【００３９】
１　音声認識装置
２　制御手段（第３の導出手段、第３の検出手段）
３　記録手段
４　記憶手段
５　音取得手段
２０　フレーム生成部
２１　スペクトルの偏倚導出部（導出手段）
２１ａ　スペクトルの偏倚／パワー／ピッチ導出部（導出手段）
２１ｂ　変化量導出部（第２の導出手段）
２２　非音声区間検出部（判定手段、検出手段）
２２ａ　非音声区間検出部（判定手段、検出手段）
２２ｂ　非音声区間検出部（判定手段、検出手段、第２の判定手段、第２の検出手段）
【発明を実施するための最良の形態】
【００４０】
以下、本発明をその実施の形態を示す図面に基づいて詳述する。
実施の形態１
図１は、本発明の実施の形態１に係る非音声区間検出装置の一実施例である音声認識装置の構成例を示すブロック図である。図中１は、例えば車両に搭載されるナビゲーション装置のようなコンピュータを用いた音声認識装置であり、音声認識装置１は、装置全体を制御するＣＰＵ（Central Processing Unit）及びＤＳＰ（Digital Signal Processor）等の制御手段２と、プログラム及びデータ等の各種情報を記録するハードディスク及びＲＯＭ等の記録手段３と、一時的に発生するデータを記録するＲＡＭからなる記憶手段４と、外部から音を取得するマイクロホンからなる音取得手段５と、音を出力するスピーカからなる音出力手段６と、液晶モニタからなる表示手段７と、目的地までの経路指示のようなナビゲーションに係る処理を実行するナビゲーション手段８とを備えている。
【００４１】
記録手段３には、本発明に係る非音声区間検出方法を実行するコンピュータプログラム３０が記録されており、記録されているコンピュータプログラム３０に含まれる各種手順を記録手段３に記憶して制御手段２の制御にて実行することにより、コンピュータは、本発明の非音声区間検出装置としても動作する。
【００４２】
また、記録手段３の記録領域の一部は、音声認識用の音響モデルを記録している音響モデルデータベース（音響モデルＤＢ）３１、音響モデルに対応する音素又は音節定義で表記された認識語彙及び文法を記録している認識辞書３２等の各種データベースとして用いられている。
【００４３】
記憶手段４の記憶領域の一部は、音取得手段５が取得したアナログ信号である音を所定の周期で標本化（サンプリング）してデジタル化した音データを記録する音データバッファ４１、及び音データを所定の時間長に区分したフレームから抽出した特徴量をはじめとするデータを記憶するフレームバッファ４２、及び一時的に発生した情報を記憶するワークメモリ４３として用いられる。
【００４４】
ナビゲーション手段８は、ＧＰＳ（Global Positioning System）のような位置検出機構と、地図情報を記録するＤＶＤ（Digital Versatile Disk）及びハードディスク等の記録媒体とを有し、現在地から目的地までの経路検索及び経路指示等のナビゲーション処理を実行し、地図及び経路を表示手段７に表示し、音声による案内を音出力手段６から出力する。
【００４５】

尚、図１に示した構成例はあくまでも一例であり、様々な形態に展開することが可能である。例えば、音声認識に係る機能を一又は複数のＶＬＳＩチップとして構成し、ナビゲーション装置に組み込むことも可能であり、音声認識用の専用装置をナビゲーション装置に外付けすることも可能である。また、制御手段２を音声認識及びナビゲーションの双方の処理で共用するようにしても、夫々専用の回路を設けるようにしてもよく、更には音声認識に関する特定の演算、例えば後述するＦＦＴ（Fast Fourier Transform）、ＤＣＴ（Discrete Cosine Transform）及びＩＤＣＴ（Inverse Discrete Cosine Transform）等の処理を実行するコプロセッサを制御手段２に組み込んでもよい。また、音データバッファ４１を音取得手段５の付属回路とし、フレームバッファ４２及びワークメモリ４３を制御手段２が備えるメモリ上に構成するようにしてもよい。更に、本発明の音声認識装置１は、ナビゲーション装置のような車載装置に限らず、音声認識を行う様々な用途の装置に用いることが可能である。

【００４６】

次に本発明の実施の形態１に係る非音声区間検出装置の一実施例である音声認識装置１の処理について説明する。図２は、制御手段２の音声認識に係る処理構成例を示すブロック図である。また、図３は、制御手段２の音声認識処理の一例を示すフローチャートである。

制御手段２は、音データからフレームを生成するフレーム生成部２０、生成されたフレームについてスペクトルの偏倚を導出するスペクトルの偏倚導出部２１、導出されたスペクトルの偏倚に基づく判定基準を用いて非音声区間を検出する非音声区間検出部２２、検出された非音声区間をもとに音声区間の開始／終了を確定させる音声区間判定部２３、及び判定された音声区間について音声を認識する音声認識部２４を備えている。

【００４７】

制御手段２は、音取得手段５によって外部の音をアナログ信号として取得し（ステップＳ１１）、取得した音を所定の周期で標本化してデジタル化した音データを、音データバッファ４１に記録する（ステップＳ１２）。ステップＳ１１にて取得する外部の音とは、人が発声する音声、定常雑音及び非定常雑音等の様々な音が重畳された音である。人が発声する音声は、音声認識装置１による認識の対象となる音声である。定常雑音は、ロードノイズ及びエンジン音等の雑音であり、既に提案及び確立されている様々な除去方法が適用される。非定常雑音としては、車両に配設されたハザード、ウインカーのようなリレー音、及びワイパーの摺動音のような機構による雑音を例示することができる。

【００４８】

そして制御手段２のフレーム生成部２０は、音データバッファ４１に記憶した音データより、１０msecのフレーム長で５msecずつオーバーラップさせたフレームを生成し（ステップＳ１３）、生成したフレームをフレームバッファ４２に記憶させる（ステップＳ１４）。尚、フレーム生成部２０は、音声認識の分野における一般的なフレーム処理として、フレーム分割前のデータに対して高域強調フィルタリング処理を施した後に、フレームに分割する。このようにして生成された各フレームに対し、以下の処理が行われる。

【００４９】

スペクトルの偏倚導出部２１は、フレーム生成部２０からフレームバッファ４２を介して与えられたフレームについて、後述するスペクトルの偏倚を導出し（ステップＳ１５）、導出したスペクトルの偏倚をフレームバッファ４２に書き込む。この場合、書き込まれたフレーム及びスペクトルの偏倚を夫々参照するのに用いられるフレームバッファ４２へのポインタ（アドレス）が、ワークメモリ４３上に設けてあり、前記ポインタを介して、フレームバッファ４２に記憶したスペクトルの偏倚にアクセスする。

尚、スペクトルの偏倚を導出する前に、ノイズキャンセル処理及びスペクトルサブトラクション処理を行って、雑音の影響を除外してもよい。

【００５０】

非音声区間検出部２２は、フレームバッファ４２を介してスペクトルの偏倚導出部２１より与えられたフレームについて、スペクトルの偏倚に基づく判定基準により非音声区間

を検出するサブルーチンを呼び出す（ステップＳ１６）。非音声区間検出部２２が判定基準を用いて検出した非音声区間のフレームは、フレームバッファ４２を介して順次音声区間判定部２３に与えられる。判定結果が未確定のフレーム、即ち後続するフレームによっては非音声区間になり得るフレームは、判定基準が用い尽くされるまで、非音声区間検出部２２によって保留される。

【００５１】

音声区間判定部２３は、非音声区間検出部２２が非音声区間として検出できなかった区間を音声区間とみなし、音声区間長が既定の最短音声区間長Ｌ１を超えた場合に音声区間開始と判定して、音声区間開始フレームを確定させる。そして音声区間が途切れたフレームを、音声区間終了点候補とする。その後、既定の最大ポーズ長Ｌ２を超えるまでに次の音声区間が始まった場合は、前述の音声区間終了点候補を棄却して、再び音声区間が途切れるのを待つ。

既定の最大ポーズ長Ｌ２を超えても次の音声区間が始まらなかった場合、音声区間判定部２３は、音声区間終了候補を音声区間終了フレームとして確定させる。音声区間の開始／終了フレームを確定したことにより、音声区間判定部２３は、一つの音声区間の判定を終える（ステップＳ１７）。このようにして検出された音声区間は、フレームバッファ４２を介して音声認識部２４に与えられる。

尚、音声区間の検出誤りを回避するため、音声区間判定部２３が判定した音声区間よりも、例えば前後に１００msecだけ広い区間を、確定させた音声区間としてもよい。

【００５２】

音声認識部２４は、音声認識の分野で一般的な技術を用いて、音声区間のフレームのデジタル信号から特徴ベクトルを抽出し、抽出した特徴ベクトルに基づいて、音響モデルデータベース３１に記録している音響モデル並びに認識辞書３２に記憶している音響語彙及び文法を参照し、入力されたフレームバッファ４２の最後（音声区間の最後）まで、音声認識処理を実行する（ステップＳ１８）。

【００５３】

図３は、一音声区間が確定した場合に、音声認識処理を実行して終了する構成であるが、音声区間を検出した場合に、計算可能なフレームから音声認識処理を実行してレスポンスタイムを短縮する構成、又は一定時間について、音声区間が検出できない場合に、処理を終了する構成としてもよい。

【００５４】

ここで、図３を用いて説明したステップＳ１５におけるスペクトルの偏倚について、更に詳述する。

本実施の例では、音データの各フレームにおけるスペクトルの傾き、即ち、スペクトルの高域／低域での偏りを示す尺度として高域・低域強度を定義する。高域・低域強度は、そのままスペクトルの偏倚として用いることができるが、本実施の例では、スペクトルの偏倚を、高域・低域強度の絶対値で表すものとする。高域・低域強度は、スペクトル包絡を近似する指標であって、音データのパワーを示す０次の自己相関関数に対する、遅れ時間が１サンプルの１次の自己相関関数の比で表すことができる。

自己相関関数は、音データを分析単位である１フレーム毎（例えば、フレーム幅：Ｎ＝256サンプル）に抽出し、ハミング窓をかけた音データの波形｛x(n)｝から、短時間自己相関関数{c(τ)}として、下記の式１より算出することができる。

【００５５】

【数１】

$$c(\tau)=\frac{1}{N-1}\sum_{n=0}^{N-2}x(n)x(n+\tau) \quad , \quad 0\leqq\tau\leqq1 \qquad \cdots 式１$$

【００５６】
また、０次及び１次の自己相関関数の比を用いるので、夫々について共通の係数である1/(N-1)を除いて、下記の式２としてもよい。
【００５７】
【数２】

$$c(\tau)=\sum_{n=0}^{N-2}x(n)x(n+\tau) \quad , \quad 0\leqq\tau\leqq1 \qquad \cdots 式２$$

【００５８】
また、自己相関関数c(τ)は、Wiener-Khintchineの定理により、短時間スペクトルS(ω)を逆フーリエ変換（ＩＤＦＴ：Inverse Discrete Fourier Transform）して算出することもできる。短時間スペクトルS(ω)は、音データを分析単位である１フレーム毎（例えば、フレーム幅：Ｎ＝256サンプル）に抽出し、各フレームに対してハミング窓をかけ、窓かけ後のフレームのデータに対してＤＦＴ（Discrete Fourier Transform）を行うことで算出できる。
尚、算出に伴う処理量を削減するため、ＩＤＦＴ／ＤＦＴに替えてＩＤＣＴ／ＤＣＴを用いることができる。
【００５９】
上述のようにして求めた自己相関関数c(τ)について、０次及び１次の比を用いて、高域・低域強度Ａを下記の式３及び式４のとおり定義する。
【００６０】
Ａ＝c(1)／ｃ(0)　(c(0)≠0)　・・・・・式３
Ａ＝0　(c(0)＝0)　・・・・・式４
【００６１】
この場合、Ａは、-1≦Ａ≦1の範囲の値をとり、１（又は-１）に近い値であるほどスペクトルの低域（又は高域）の強度が大きいことを示す。
尚、高域・低域強度としては、上述したＡに限定されるものではなく、０次及び１次以外の異なる次数についての自己相関関数の比、所定周波数帯域のパワー、所定の異なる周波数帯域についてのパワーの比、ＭＦＣＣ、対数スペクトラムを逆フーリエ変換したケプストラム、又は推定したフォルマントのうち所定の異なるフォルマントについての周波数の比若しくはパワーの比の少なくとも一であってもよい。複数の高域・低域強度を導出した場合は、夫々導出した値に基づいて、非音声区間の判定を並列的に実行することができる。
【００６２】
図５乃至８は、夫々鼻をすする音、踏切の警報音及び２種類の発声音（「えーテスト中です」、「経営（けーえー）」）について、パワー及び高域・低域強度等のデータを示す図である。図５乃至８の各図において、横軸は時間であり、縦軸は、上から音データの波形、音データのパワー（鎖線、左軸）、高域・低域強度Ａ（実線、右軸）及びスペクトログラム（左軸）である。

【００６３】
図５では、スペクトログラムにおいて、黒の濃い領域が高域である上方に偏っているため、当該区間でＡの値は－１に近づいている。
図６では、警報のトーン信号により、スペクトログラムの下半分に黒の濃い線が出現して、低域に偏っているため。Ａの値は１に近づいている。
【００６４】
図７では、発声されている音素によって、高域／低域が強い、又はどちらでもない、という区間が出現しており、Ａの値は概ね-０．７＜Ａ＜０．７の範囲で大きく変動している。即ち、発声中の区間では、Ａの値は長時間特定の値に留まることがなく、ある程度の範囲で変動するといえる。発声中であってもＡの値が安定するのは、図７の発声末尾の「す」のように、同じ音素が継続している場合である。この場合、「す」が無声化して、高域が強い摩擦音/s/が継続しているため、Ａの値は－１に近い－０．７近辺で約０．３秒間に渡り安定している。また、同じように１音素が継続する区間であっても、発声される音素によってＡの値は変動する。例えば、図７では、「テスト中」末尾の「う」近辺で、母音/u/が継続しているが、Ａの値はプラス方向に振れ、０．６前後の値をとっている。
【００６５】
一方、日本語の語彙においては、特定の母音／子音が無意味に連なることはないため、一般的な音声認識処理では、一つの音素が長時間発声されることは考慮する必要がない。このため、一般の単語又は文の発声において各音素が継続され得る時間長と、各音素の発声においてＡの値が取り得る範囲とを想定することにより、音素が想定外に継続した場合、又はＡの値が想定外となった場合は、当該単語又は文は音声でないと見做すことができる。例えば、図８では、「経営」を「けーえー」と発声する場合があり、最初の/k/以外は、/e/が約４モーラ長だけ継続する。この場合は、日本語において同一の音素が最も長時間継続する場合と想定され、その継続時間は、ゆっくりと発声された場合であっても高々1.2秒程度である。
【００６６】
上述した内容及び図５乃至８に示された事項より、スペクトルの偏倚|Ａ|について、例えば音声区間では、|Ａ|≧０．７とはならないこと、また、音素は高々１．２秒しか継続せず、当該区間で|Ａ|≧０．５とならないことがいえるため、非音声区間について、例えば下記のような判定を行うことが可能である。
（ａ）：|Ａ|≧０．７が０．１秒以上継続する場合、当該区間は非音声とする。
（ｂ）：|Ａ|≧０．５が１．２秒以上継続する場合、当該区間は非音声とする。
また、上記の判定を更に細分化して、以下のような判定を行うことも可能である。
（ｃ）：|Ａ|≧０．６が０．５秒以上継続する場合、当該区間は非音声とする。
尚、フレームが継続する時間に係る閾値は、フレーム長が一定であるため、フレームが継続する数に係る閾値に置き換えることができる。また、音取得手段５のマイクロホンの特性を含む音入力系の伝達特性によっては、高域・低域のバランスが変動してスペクトルの偏倚|Ａ|も変化することが想定されるため、入力系の伝達特性に応じて上述した判定の閾値を調整することが望ましい。
【００６７】
上述した内容を踏まえて、非音声区間検出のサブルーチンについて説明する。図４は、非音声区間検出のサブルーチンに係る制御手段２の処理手順を示すフローチャートである。非音声区間検出のサブルーチンが呼び出された場合、制御手段２は、そのときのポインタが示すフレームのスペクトルの偏倚が、所定の閾値（例えば上述した０．７）以上であるか否かを判定する（ステップＳ２１）。所定の閾値未満であると判定した場合（ステップＳ２１：ＮＯ）、制御手段２は、ワークメモリ４３に記憶されたフレームバッファ４２へのポインタを１フレーム後方に更新して（ステップＳ２２）、リターンする。
これにより、制御手段２は、非音声区間を検出することなくリターンする。
【００６８】
所定の閾値以上であると判定した場合（ステップＳ２１：ＹＥＳ）、制御手段２は、そ

のときのポインタが示すフレームのフレーム番号を「開始フレーム番号」としてワークメモリ４３上に記憶する（ステップＳ２３）。そして、制御手段２は、ワークメモリ４３上に設けた「フレームカウント」の記憶値を「１」に初期化する（ステップＳ２４）。ここで、「フレームカウント」は、スペクトルの偏倚と所定の閾値との比較判定を行ったフレーム数を計数するものである。

【００６９】

その後、制御手段２は、「フレームカウント」の記憶内容が所定数（例えば上述した0.1秒間に含まれるフレームの数である10）以上であるか否かを判定し（ステップＳ２５）、所定数未満であると判定した場合（ステップＳ２５：ＮＯ）、制御手段２は、「フレームカウント」の記憶内容に「１」を加算すると共に（ステップＳ２６）、フレームバッファへのポインタを１フレーム後方に更新する（ステップＳ２７）。そして、制御手段２は、そのときのポインタが示すフレームのスペクトルの偏倚が、所定の閾値以上であるか否かを判定する（ステップＳ２８）。

【００７０】

スペクトルの偏倚が所定の閾値以上であると判定した場合（ステップＳ２８：ＹＥＳ）、制御手段２は、処理をステップＳ２５に戻す。

スペクトルの偏倚が所定の閾値未満であると判定した場合（ステップＳ２８：ＮＯ）、制御手段２は、「開始フレーム番号」の内容を消去して（ステップＳ２９）、リターンする。

これにより、制御手段２は、非音声区間を検出することなくリターンする。

【００７１】

ステップＳ２５で「フレームカウント」の記憶内容が所定数以上であると判定した場合（ステップＳ２５：ＹＥＳ）、制御手段２は、非音声区間の終了フレームを検出する処理に移り、フレームバッファへのポインタを１フレーム後方に更新する（ステップＳ３０）。そして、制御手段２は、そのときのポインタが示すフレームのスペクトルの偏倚が、所定の閾値以上であるか否かを判定する（ステップＳ３１）。

【００７２】

スペクトルの偏倚が所定の閾値以上であると判定した場合（ステップＳ３１：ＹＥＳ）、制御手段２は、処理をステップＳ３０に戻す。スペクトルの偏倚が所定の閾値未満であると判定した場合（ステップＳ３１：ＮＯ）、制御手段２は、そのときのポインタが示すフレームの１つ前のフレーム番号を「終了フレーム番号」としてワークメモリ４３上に記憶し（ステップＳ３２）、リターンする。

これにより、「開始フレーム番号」及び「終了フレーム番号」で区切られた区間が、検出された非音声区間となる。

【００７３】

このように、本発明の実施の形態１では、各フレームの音データより導出したスペクトルの偏倚|Ａ|が、例えば０．７以上となるフレームが、継続時間にして０．１秒に相当する数以上連なる場合、スペクトルの偏倚が最初に０．７以上となったフレームから、最後に０．７以上となったフレームまでを非音声区間として検出する。

これにより、本実施の形態１では、スペクトルの偏倚が大きくて非音声の特徴を有するフレームが、音声らしからぬ程度まで連なる区間を非音声区間として検出し、人の発声による基準値の補正を要しない。従って、パワーの大きい雑音、又は非定常性の強い雑音が発生する環境下においても、発声前後か否かに拘わらず、高精度に非音声区間を検出することが可能である。

【００７４】

実施の形態２

実施の形態２は、推定背景雑音パワーを基本とした音声区間検出装置と、実施の形態１に係る非音声区間検出装置とを併用した形態である。

図９は、本発明の実施の形態２に係る非音声区間検出装置の一実施例である音声認識装置１について、制御手段２の音声認識に係る処理構成例を示すブロック図である。

【００７５】
制御手段２は、フレーム生成部２０、スペクトルの偏倚導出部２１、導出されたスペクトルの偏倚に基づく判定基準を用いて非音声区間を検出する非音声区間検出部２２ａ、検出された非音声区間をもとに音声区間の開始／終了を確定させる音声区間判定部２３ａ、確定された音声区間について音声認識の照合に用いる特徴量を算出する特徴量算出部２８、及び算出された特徴量を用いて音声認識のための照合処理を行う照合部２９を備えている。
制御手段２は、更に、フレーム生成部２０で生成されたフレームについて、音データのパワーを導出するパワー導出部２６、導出したパワーに基づいて背景雑音パワーを推定する背景雑音パワー推定部２７、及び音声区間判定部２３ａに修正すべきフレーム番号を通知する音声区間修正部２５を備える。
【００７６】
非音声区間検出部２２ａは、検出した非音声区間のフレーム番号を音声区間判定部２３ａ及び音声区間修正部２５に与える。
音声区間修正部２５は、非音声区間検出部２２ａが非音声区間として検出したフレームが、音声区間判定部２３ａでは音声区間と判定されていた場合に、音声区間判定部２３ａに対して、所定の修正信号及び修正すべきフレーム番号を与える。
【００７７】
パワー導出部２６は、フレーム生成部２０から与えられた各フレームについて音データのパワーを導出し、導出したパワーを背景雑音パワー推定部２７に与える。
尚、パワーを算出する前に、ノイズキャンセル処理及びスペクトルサブトラクション処理を行って、雑音の影響を除外してもよい。
【００７８】
背景雑音パワー推定部２７は、音データの先頭フレームを無条件に雑音とみなし、当該フレームの音データのパワーを推定背景雑音パワーの初期値とする。その後、背景雑音パワー推定部２７は、音声区間判定部２３ａから通知された音声区間のフレームを除いて、音データの２フレーム目以降について、直近の２フレームのパワーの単純移動平均をとり、導出した移動平均値によって推定背景雑音パワーをフレーム毎に更新する。尚、推定背景雑音パワーの更新値を、パワーの単純移動平均から導出するのではなく、ＩＩＲ（Infinite Impulse Response）フィルタによって導出するようにしてもよい。
また、背景雑音パワー推定部２７は、音声区間判定部２３ａより後述する推定背景雑音パワーの修正を通知された場合、非音声区間に修正されたフレームのうち、その時の最新のフレームの音データから導出されたパワーにより、推定背景雑音パワーを上書きして修正する。
【００７９】
尚、背景雑音パワー推定部２７は、音声区間判定部２３ａより推定背景雑音パワーの修正を通知された場合、非音声区間に修正されたフレームの音データについて、推定背景雑音パワーを導出するようにしてもよい。また、所定のＮ回目（Ｎは２以上の自然数）の修正を通知された場合に初めて、その時の最新のフレームの音データから導出されたパワーにより、推定背景雑音パワーを上書きするようにしてもよい。これにより、背景雑音レベルが上下に変動した場合に、推定背景雑音レベルが上がり過ぎて音声区間が検出できなくなるのを防止することができる。
【００８０】
音声区間判定部２３ａは、各フレームの音データのパワーが、「推定背景雑音パワー＋所定の閾値α」以上となった場合、当該フレームを音声区間と判定する。また、音声区間判定部２３ａは、音声区間修正部２５より上述した所定の修正信号を与えられた場合、修正すべきフレーム番号に基づいて、音声区間の判定結果を修正する。そして、音声区間判定部２３ａは、判定した音声区間が最短入力時間長以上、且つ最長入力時間長以下だけ継続した場合、その時の音声区間を確定させ、確定させた音声区間を特徴量算出部２８、照合部２９及び背景雑音パワー推定部２７に通知する。

更に、音声区間判定部２３ａは、背景雑音パワー推定部２７に対し、非音声区間に修正されたフレームの音データにより、推定背景雑音パワーを修正するように通知する。

【００８１】

特徴量算出部２８は、音声区間判定部２３ａが最終的に音声区間と確定させた区間について、音声認識の照合に用いる特徴量を算出する。ここでの特徴量とは、例えば音響モデルデータベース３１に記録している音響モデルとの類似度計算が可能な特徴ベクトルであり、フレーム処理されたデジタル信号を変換することにより導出される。本実施の形態における特徴量はＭＦＣＣ（Mel Frequency Cepstrum Coefficient）であるが、ＬＰＣ（Linear Predictive Coding）ケプストラム又はＬＰＣ係数であってもよい。ＭＦＣＣは、フレーム処理されたデジタル信号をＦＦＴにて変換し、振幅スペクトルを求め、中心周波数がメル周波数領域で一定間隔であるメルフィルタバンクにて処理し、処理の結果の対数をＤＣＴにて変換し、１次乃至１４次等の低次の係数をＭＦＣＣと呼ばれる特徴ベクトルとして用いる。尚、次数については、標本化周波数及びアプリケーション等の要因により決定され、数値は限定されない。

【００８２】

照合部２９は、音声区間判定部２３ａが音声と判定し確定させた音声区間について、特徴量算出部２８が導出した特徴量である特徴ベクトルに基づいて、音響モデルデータベース３１に記録している音響モデル並びに認識辞書３２に記録している認識語彙及び文法を参照し、音声認識処理を実行する。また、認識結果に基づいて、音出力手段６及び表示手段７等の他の入出力手段に対して出力を制御する。

【００８３】

その他、実施の形態１に対応する部分には同一符号を付して、それらの説明を省略する。

【００８４】

このように、本発明の実施の形態２では、音データのパワーを基本とした音声区間検出装置の検出結果を、本発明に係る非音声区間検出装置により修正することが可能となり、全体として音声区間検出の精度を向上させることができる。

【００８５】

実施の形態３

実施の形態３は、実施の形態１及び２でスペクトルの偏倚に基づいて非音声区間を検出するのに対し、スペクトルの偏倚、音データのパワー又は音データのピッチについての前フレームとの変化量に基づいて、非音声区間を検出する形態である。また、非音声区間の検出対象から除外する区間を検出し、更に検出対象から除外された区間を復活させる処理をも含む形態である。図１０は、本発明の実施の形態３に係る非音声区間検出装置の一実施例である音声認識装置１について、制御手段２の音声認識に係る処理構成例を示すブロック図である。また、図１１は、制御手段２の音声認識処理の一例を示すフローチャートである。

【００８６】

制御手段２は、音データからフレームを生成するフレーム生成部２０、生成されたフレームについて、音データのスペクトルの偏倚／パワー／ピッチを導出するスペクトルの偏倚／パワー／ピッチ導出部２１ａ、導出されたスペクトルの偏倚／パワー／ピッチについて前フレームとの変化量を導出する変化量導出部２１ｂ、導出された変化量に基づく判定基準を用いて非音声区間を検出する非音声区間検出部２２ｂ、検出された非音声区間をもとに音声区間の開始／終了を確定させる音声区間判定部２３ｂ、及び判定された音声区間について音声を認識する音声認識部２４を備えている。

【００８７】

ステップＳ４１乃至Ｓ４４の処理は、夫々図３のステップＳ１１乃至Ｓ１４と同様であるので、説明を省略する。ステップＳ４１乃至Ｓ４４の処理で生成された各フレームに対し、以下の処理が行われる。

【００８８】

スペクトルの偏倚／パワー／ピッチ導出部２１ａは、フレーム生成部２０からフレームバッファ４２を介して与えられたフレームについて、音データのスペクトルの偏倚、音データのパワー及び音データのピッチの少なくとも一を導出し（ステップＳ４５）、導出したスペクトルの偏倚、パワー及びピッチの少なくとも一をフレームバッファ４２に書き込む。

尚、ここで導出する値は、スカラー量であるスペクトルの偏倚／パワー／ピッチに限定されるものではなく、音響的な特性を表すベクトルであるパワースペクトル、振幅スペクトル、ＭＦＣＣ、ＬＰＣケプストラム、ＬＰＣ係数、ＰＬＰ係数又はＬＳＰパラメータであってもよい。

【００８９】

変化量導出部２１ｂは、フレームバッファ４２に書き込まれたスペクトルの偏倚、音データのパワー及び音データのピッチの少なくとも一について、前フレームとの変化量を導出してフレームバッファ４２に書き込む（ステップＳ４６）。この場合、書き込まれたフレーム及び変化量を夫々参照するのに用いられるフレームバッファ４２へのポインタ（アドレス）が、ワークメモリ４３上に設けられ、初期化される。

【００９０】

非音声区間検出部２２ｂは、フレームバッファ４２を介して変化量導出部２１ｂより与えられたフレームについて、変化量に基づく判定基準により非音声区間を検出するサブルーチンを呼び出す（ステップＳ４７）。非音声区間検出部２２ｂが判定基準を用いて検出した非音声区間のフレームは、フレームバッファ４２を介して順次音声区間判定部２３ｂに与えられる。その後、音声区間判定部２３ｂは、音声区間の開始／終了フレームを確定して音声区間の判定を行う（ステップＳ４８）。そして、音声認識部２４は、入力されたフレームバッファ４２の最後（音声区間の最後）まで、音声認識処理を実行する（ステップＳ４９）。

【００９１】

ここで、図１１を用いて説明したステップＳ４６における変化量について、更に詳述する。

人が発声した場合の音データは、スペクトルの偏倚、パワー及びピッチの何れについても、時間と共にある程度の変動が生じるのを避けられない。逆に音データの上記指標に変動が観測されない場合は、非音声であると見做すのが適当である。

例えば、ｔ番目のフレーム（以下、フレームｔという。ｔ＝１、２、・・・）における高域・低域強度ＡをＡ(t)とするとき、フレームｔでの変化量を下記の式５及び式６のとおり定義する。

【００９２】

$C(t)=|A(t)-A(t-1)|,\quad t>1$　　・・・・式５

$C(t)=0,\quad t=1$　　・・・・式６

【００９３】

この場合、非音声区間について、例えば下記のような判定を行うことが可能である。

（ｄ）：C(t)≦０．０５のフレームが０．５秒以上継続する場合は、非音声とする。

（ｅ）：C(t)≦０．１のフレームが１．２秒以上継続する場合は、非音声とする。

【００９４】

尚、C(t)による判定は、上記（ｄ）、（ｅ）に限定されるものではなく、変化量に係る閾値と継続時間に係る閾値との組み合わせにより、異なる条件を設定することが可能である。また、フレームが継続する時間に係る閾値は、フレーム長が一定であるため、フレームの継続する数に係る閾値に置き換えることができる。

更に、スペクトルの偏倚、音データのパワー及び音データのピッチ夫々について変化量を別々に導出し、夫々の変化量について、図１１のステップＳ４７を実行して、非音声区間を別々に検出することも可能である。

【００９５】

一方、上述の（ｄ）、（ｅ）の判定基準とは逆に、変化量が大きいフレームは非音声で

ない可能性があるため、例えば下記（ｆ）の判定を加えることが有効である。
（ｆ）：Ｃ(t)＞０．５の場合、ｔ－ｗ＋１（例えばｗ＝３）からｔ＋ｗ-１のフレームを非音声区間の検出対象から除外する。即ちそのときのフレームを含めて前後にｗだけ連なるフレームからなる区間を、非音声区間の検出対象から除外する。

【００９６】

また、上記（ｆ）の判定に拘わらず、変化量が大きいフレームが連なる区間が所定数より短い場合は、単発的に変化量が増大した非音声区間である可能性があるため、例えば下記（ｇ）の判定を更に加えることが望ましい。
（ｇ）：（ｆ）により、変化量が大きいと判定されるフレームが連なる数が所定数以下であって、（ｆ）により非音声区間の検出対象から除外されている区間が、非音声区間に挟まれている場合は、（ｆ）の判定を覆して非音声区間として検出する。

【００９７】

上述した内容を踏まえて、非音声区間検出のサブルーチンについて説明する。図１２は、非音声区間検出のサブルーチンに係る制御手段２の処理手順を示すフローチャートである。非音声区間検出のサブルーチンが呼び出された場合、制御手段２は、そのときのポインタが示すフレームの変化量が、所定の閾値（例えば上述した０．０５）以下であるか否かを判定する（ステップＳ５１）。所定の閾値以下であると判定した場合（ステップＳ５１：ＹＥＳ）、制御手段２は、非音声区間検出確定のサブルーチンを呼び出し（ステップＳ５２）、その後リターンする。

【００９８】

変化量が所定の閾値を超えると判定した場合（ステップＳ５１：ＮＯ）、制御手段２は、変化量が第２の閾値（例えば上述した0.5）を超えるか否かを判定する（ステップＳ５３）。第２の閾値を超えないと判定した場合（ステップＳ５３：ＮＯ）、制御手段２はそのままリターンする。

変化量が第２の閾値を超えると判定した場合（ステップＳ５３：ＹＥＳ）、制御手段２は、非音声区間検出除外のサブルーチンを呼び出し（ステップＳ５４）、その後リターンする。

【００９９】

図１３及び図１４は、非音声区間検出除外のサブルーチンに係る制御手段２の処理手順を示すフローチャートであり、図１５及び図１６は、非音声区間検出確定のサブルーチンに係る制御手段２の処理手順を示すフローチャートである。図１３及び図１４について、非音声区間検出除外のサブルーチンが呼び出された場合、制御手段２は、そのときのポインタが示すフレームのフレーム番号を「開始フレーム番号」としてワークメモリ４３上に記憶する（ステップＳ６１）。そして、制御手段２は、ワークメモリ４３上に設けた「フレームカウント」の記憶値を「１」に初期化する（ステップＳ６２）。ここで、「フレームカウント」は、変化量と第２の閾値との比較判定を行ったフレーム数を計数するものである。

【０１００】

その後、制御手段２は、「フレームカウント」の記憶内容が所定数（例えば３０msecの間に含まれるフレームの数である３）以下であるか否かを判定し（ステップＳ６３）、所定数以下であると判定した場合（ステップＳ６３：ＹＥＳ）、制御手段２は、「フレームカウント」の記憶内容に「１」を加算すると共に（ステップＳ６４）、フレームバッファへのポインタを１フレーム後方に更新する（ステップＳ６５）。そして、制御手段２は、そのときのポインタが示すフレームの変化量が、上述した所定の閾値より大きい第２の閾値を超えるか否かを判定する（ステップＳ６６）。

【０１０１】

変化量が第２の閾値を超えると判定した場合（ステップＳ６６：ＹＥＳ）、制御手段２は、処理をステップＳ６３に戻す。変化量が第２の閾値以下であると判定した場合（ステップＳ６６：ＮＯ）、即ち単発的に変化量が増大した区間が終了した場合、制御手段２は、「開始フレーム番号」に記憶しているフレームに対して「第２の所定数」フレーム前（

ここでは、上述のｗフレーム前）が、非音声区間であるか否かを判定する（ステップＳ６７）。「第２の所定数」フレーム前が非音声区間であると判定した場合（ステップＳ６７：ＹＥＳ）、制御手段２は、単発的に変化量が増大した区間が、後に非音声区間と判定される可能性があるものとして、当該区間に「非音声候補区間」のマークを付与する（ステップＳ６８）。

【０１０２】

ステップＳ６３で「フレームカウント」の記憶内容が所定数を超えると判定した場合（ステップＳ６３：ＮＯ）、即ち、変化量の大きい区間が単発的とは言えない程度に継続した場合、制御手段２は、当該区間の終了フレームを検出する処理に移り、フレームバッファへのポインタを１フレーム後方に更新する（ステップＳ６９）。そして、制御手段２は、そのときのポインタが示すフレームの変化量が、第２の閾値を超えるか否かを判定する（ステップＳ７０）。変化量が第２の閾値を超えると判定した場合（ステップＳ７０：ＹＥＳ）、制御手段２は、処理をステップＳ６９に戻す。

【０１０３】

変化量が第２の閾値以下であると判定した場合（ステップＳ７０：ＮＯ）、即ち変化量が第２の閾値より増大した区間が終了した場合、又はステップＳ６７で「第２の所定数」フレーム前が非音声区間でないと判定した場合（ステップＳ６７：ＮＯ）、制御手段２は、変化量が増大した区間を非音声区間の検出対象から除外するために、当該区間に「非音声除外区間」のマークを付与する（ステップＳ７１）。

【０１０４】

ステップＳ７１の処理を終えた場合、又はステップＳ６８の処理を終えた場合、制御手段２は、「開始フレーム番号」の内容から「第２の所定数（ここでは上述のｗ）-１」を減じる処理を行う（ステップＳ７２）。更に、制御手段２は、そのときのポインタが示すフレームの１つ前のフレーム番号に「第２の所定数（ここでは上述のｗ）-１」を加えた数を「終了フレーム番号」としてワークメモリ４３上に記憶し（ステップＳ７３）、リターンする。

これにより、変化量が第２の閾値を超えた区間を、前後に「ｗ-１」だけ拡張した区間が、「非音声候補区間」又は「非音声除外区間」の扱いとなる。

【０１０５】

次に、図１５及び図１６について、非音声区間検出確定のサブルーチンが呼び出された場合、制御手段２は、そのときのポインタが示すフレームのフレーム番号を「開始フレーム番号」としてワークメモリ４３上に記憶する（ステップＳ８１）。そして、制御手段２は、ワークメモリ４３上に設けた「フレームカウント」の記憶値を「１」に初期化する（ステップＳ８２）。ここで、「フレームカウント」は、変化量と所定の閾値との比較判定を行ったフレーム数を計数するものである。

【０１０６】

その後、制御手段２は、「フレームカウント」の記憶内容が、ステップS６３での所定数とは異なる所定数（例えば上述の０．５秒の間に含まれるフレームの数）以上であるか否かを判定し（ステップＳ８３）、所定数未満であると判定した場合（ステップＳ８３：ＮＯ）、制御手段２は、「フレームカウント」の記憶内容に「１」を加算すると共に（ステップＳ８４）、フレームバッファへのポインタを１フレーム後方に更新する（ステップＳ８５）。そして、制御手段２は、そのときのポインタが示すフレームの変化量が、所定の閾値以下であるか否かを判定する（ステップＳ８６）。

【０１０７】

変化量が所定の閾値以下であると判定した場合（ステップＳ８６：ＹＥＳ）、制御手段２は、処理をステップＳ８３に戻す。変化量が所定の閾値を超えると判定した場合（ステップＳ８６：ＮＯ）、即ち変化量が所定の閾値以下であるフレームが所定数未満しか継続しなかった場合、制御手段２は、非音声区間を検出しなかったものとし、「開始フレーム番号」に記憶したフレームの直前のフレームが、非音声候補区間に含まれるか否かを判定する（ステップＳ８７）。

【０１０８】
直前のフレームが非音声候補区間に含まれていると判定した場合（ステップＳ８７：ＹＥＳ）、制御手段２は、当該非音声候補区間を非音声除外区間に変更する（ステップＳ８８）。直前のフレームが非音声候補区間に含まれていないと判定した場合（ステップＳ８７：ＮＯ）、又はステップＳ８８の処理を終えた場合、制御手段２は、「開始フレーム番号」の記憶内容を消去して（ステップＳ８９）、リターンする。
【０１０９】
ステップＳ８３で「フレームカウント」の記憶内容が所定数以上であると判定した場合（ステップＳ８３：ＹＥＳ）、制御手段２は、非音声区間の終了フレームを検出する処理に移り、フレームバッファへのポインタを１フレーム後方に更新する（ステップＳ９０）。そして、制御手段２は、そのときのポインタが示すフレームの変化量が、所定の閾値以下であるか否かを判定する（ステップＳ９１）。変化量が所定の閾値以下であると判定した場合（ステップＳ９１：ＹＥＳ）、制御手段２は、処理をステップＳ９０に戻す。
【０１１０】
変化量が所定の閾値を超えると判定した場合（ステップＳ９１：ＮＯ）、即ち検出した非音声区間が終了した場合、制御手段２は、「開始フレーム番号」に記憶したフレームの直前のフレームが、非音声候補区間に含まれるか否かを判定する（ステップＳ９２）。直前のフレームが非音声候補区間に含まれていると判定した場合（ステップＳ９２：ＹＥＳ）、制御手段２は、当該非音声候補区間のマークを消去して、非音声区間に確定させる（ステップＳ９３）。
【０１１１】
直前のフレームが非音声候補区間に含まれていないと判定した場合（ステップＳ９２：ＮＯ）、又はステップＳ９３の処理を終えた場合、制御手段２は、そのときのポインタが示すフレームの１つ前のフレーム番号を「終了フレーム番号」としてワークメモリ４３上に記憶し（ステップＳ９４）、リターンする。
これにより、「開始フレーム番号」及び「終了フレーム番号」で区切られた区間が、新たに検出された非音声区間となる。
【０１１２】
その他、実施の形態１又は２に対応する部分には同一符号を付して、それらの説明を省略する。
【０１１３】
このように、本発明の実施の形態３では、各フレームの音データより導出したスペクトルの偏倚、パワー及びピッチの少なくとも一について、前フレームとの変化量Ｃ(t)が、例えば０．０５以下となるフレームが、継続時間にして０．５秒に相当する数以上連なる場合、変化量が最初に０．０５以下となったフレームから、最後に０．０５以下となったフレームまでを非音声区間として検出する。また、単発的に変化量の大きい区間は非音声区間の検出対象から除外し、更に当該区間が非音声区間に挟まれている場合は、判定を覆して非音声区間として検出する。
これにより、本実施の形態３では、変化量が小さくて非音声の特徴を有するフレームが、音声らしからぬ程度まで連なる区間を非音声区間として検出し、人の発声による基準値の補正を要しない。従って、パワー変動の大きい雑音が発生する環境下においても、発声前後か否かに拘わらず、高精度に非音声区間を検出することが可能である。また、単発的に変化量が大きい区間（例えば、エアコンの風量が変動して、定量的な雑音が変化した瞬間）についても、適切に非音声区間の検出を行うことが可能となる。
【０１１４】
尚、実施の形態３にあっては、変化量導出部２１ｂがフレームｔにおいて導出する変化量Ｃ(t)は、上述の式５及び式６に限定されるものではなく、フレームｔの前後ｖ（例えばｖ＝２）フレームの区間、即ちフレームｔ－ｖからフレームｔ＋ｖの区間において、下記の式７又は式８で定義される最大値であってもよい。
【０１１５】

【数３】

$$D(t)=\max_{j\leqq i\leqq t+v}A(i)-\min_{j\leqq i\leqq t+v}A(i)\quad,\quad j=\max(0,\quad t-v)\qquad\cdots\text{式７}$$

【数４】

$$E(t)=\max_{j\leqq i\leqq t+v}C(i)\quad,\quad j=\max(0,\quad t-v)\qquad\cdots\text{式８}$$

【０１１６】
これにより、変化量はＣ(t)近傍のフレームにおける変化量の最大値と置き換わるため、非音声区間が検出され難くなって、非音声区間を誤検出することを抑止することができる。
【０１１７】
また、実施の形態１（又は実施の形態３）にあっては、スペクトルの偏倚導出部２１（又はスペクトルの偏倚／パワー／ピッチ導出部２１ａ）は、フレームｔの前後ｚ（例えばｚ＝３）フレームの区間、即ちフレームｔ－ｚからフレームｔ＋ｚの区間におけるスペクトルの偏倚の最大値、最小値、平均値及び中央値の少なくとも一を導出して、導出した値を夫々フレームｔについてのスペクトルの偏倚としてもよい。これらの統計的な集計値を用いることにより、短時間で急激な信号変化があった場合に、スペクトルの偏倚の誤認識を防止することができる。この場合、新たに導出した夫々のスペクトルの偏倚について、非音声区間を別々に検出することが可能である。
【０１１８】
実施の形態４
実施の形態４は、実施の形態１において、スペクトルの偏倚が所定の閾値以上となるフレームが、所定数以上連なる区間を非音声区間として検出するのに対し、スペクトルの偏倚が所定の閾値以上となるフレームが、所定の割合を超える区間について、当該区間が所定数以上のフレームに亘って連なる場合、当該区間を非音声区間として検出する形態である。
図１７及び図１８は、本発明の実施の形態４に係る非音声検出装置の一実施例である音声認識装置１について、非音声区間検出のサブルーチンに係る制御手段２の処理手順を示すフローチャートである。
【０１１９】
非音声区間検出のサブルーチンが呼び出された場合、制御手段２は、そのときのポインタが示すフレームのスペクトルの偏倚が、所定の閾値以上であるか否かを判定する（ステップＳ１１１）。所定の閾値未満であると判定した場合（ステップＳ１１１：ＮＯ）、制御手段２は、ワークメモリ４３に記憶されたフレームバッファ４２へのポインタを１フレーム後方に更新して（ステップＳ１１２）、リターンする。
これにより、制御手段２は、非音声区間を検出することなくリターンする。
【０１２０】
所定の閾値以上であると判定した場合（ステップＳ１１１：ＹＥＳ）、制御手段２は、そのときのポインタが示すフレームのフレーム番号を「開始フレーム番号」としてワークメモリ４３上に記憶する（ステップＳ１１３）。そして、制御手段２は、ワークメモリ４３上に設けた「フレームカウント１」の記憶値を「１」に初期化し（ステップＳ１１４）、更に「フレームカウント２」の記憶値を「１」に初期化する（ステップＳ１１５）。ここで、「フレームカウント１」は、スペクトルの偏倚と所定の閾値との比較判定を行ったフレーム数を計数するものである。また、「フレームカウント２」は、スペクトルの偏倚

が所定の閾値以上となったフレーム数を計数するものである。

【０１２１】

その後、制御手段２は、「フレームカウント１」の記憶内容が所定数以上であるか否かを判定し（ステップＳ１１６）、所定数未満であると判定した場合（ステップＳ１１６：ＮＯ）、制御手段２は、「フレームカウント１」の記憶内容に「１」を加算すると共に（ステップＳ１１７）、フレームバッファへのポインタを１フレーム後方に更新する（ステップＳ１１８）。そして、制御手段２は、そのときのポインタが示すフレームのスペクトルの偏倚が、所定の閾値以上であるか否かを判定する（ステップＳ１１９）。

【０１２２】

スペクトルの偏倚が所定の閾値以上であると判定した場合（ステップＳ１１９：ＹＥＳ）、制御手段２は、「フレームカウント２」の記憶内容に「１」を加算して（ステップＳ１２０）、処理をステップＳ１１６に戻す。スペクトルの偏倚が所定の閾値未満であると判定した場合（ステップＳ１１９：ＮＯ）、制御手段２は、「フレームカウント１」の記憶内容に対する「フレームカウント２」の記憶内容の比、即ちスペクトルの偏倚を判定した全フレームに対する、スペクトルの偏倚が所定の閾値以上となったフレームの割合が、所定の割合（例えば０．８）以上であるか否かを判定する（ステップＳ１２１）。

【０１２３】

所定の割合以上であると判定した場合（ステップＳ１２１：ＹＥＳ）、制御手段２は、処理をステップＳ１１６に戻す。所定の割合未満であると判定した場合（ステップＳ１２１：ＮＯ）、制御手段２は、「開始フレーム番号」の内容を消去して（ステップＳ１２２）、リターンする。

これにより、制御手段２は、非音声区間を検出することなくリターンする。

【０１２４】

ステップＳ１１６で「フレームカウント１」の記憶内容が所定数以上であると判定した場合（ステップＳ１１６：ＹＥＳ）、制御手段２は、非音声区間の終了フレームを検出する処理に移り、「フレームカウント」の記憶内容に「１」を加算すると共に（ステップＳ１２３）、フレームバッファへのポインタを１フレーム後方に更新する（ステップＳ１２４）。そして、制御手段２は、そのときのポインタが示すフレームのスペクトルの偏倚が、所定の閾値以上であるか否かを判定する（ステップＳ１２５）。

【０１２５】

スペクトルの偏倚が所定の閾値以上であると判定した場合（ステップＳ１２５：ＹＥＳ）、制御手段２は、「フレームカウント２」の記憶内容に「１」を加算する（ステップＳ１２６）。ステップＳ１２６の処理を終えた場合、又はスペクトルの偏倚が所定の閾値未満であると判定した場合（ステップＳ１２５：ＮＯ）、制御手段２は、「フレームカウント１」の記憶内容に対する「フレームカウント２」の記憶内容の比が、所定の割合以上であるか否かを判定する（ステップＳ１２７）。

【０１２６】

所定の割合以上であると判定した場合（ステップＳ１２７：ＹＥＳ）、制御手段２は、処理をステップＳ１２３に戻す。所定の割合未満であると判定した場合（ステップＳ１２７：ＮＯ）、制御手段２は、そのときのポインタが示すフレームの１つ前のフレーム番号を「終了フレーム番号」としてワークメモリ４３上に記憶し（ステップＳ１２８）、リターンする。

これにより、「開始フレーム番号」及び「終了フレーム番号」で区切られた区間が、検出された非音声区間となる。

【０１２７】

その他、実施の形態１に対応する部分には同一符号を付して、それらの説明を省略する。

【０１２８】

このように、本発明の実施の形態４では、各フレームの音データより導出したスペクトルの偏倚が所定の閾値以上となるフレームが、所定の割合を超える区間について、当該区

間が所定数以上のフレームに亘って連なる場合、スペクトルの偏倚が最初に所定の閾値以上となったフレームから、スペクトルの偏倚が所定の閾値以上となるフレームの割合が所定の割合未満となる直前のフレームまでを非音声区間として検出する。

これにより、スペクトルの偏倚が、短時間に変動する場合であっても、高精度に非音声区間を検出することができる。

【０１２９】

尚、検出する非音声区間の先頭フレームは、最初に所定の閾値以上となったフレームに限定されず、スペクトルの偏倚が所定の閾値以上となるフレームの割合が所定の割合以上である範囲において、前方のフレームまで遡ったフレームを先頭フレームとしてもよい。

【０１３０】

実施の形態５

実施の形態５は、実施の形態１に対し、信号対雑音比を導出し、導出した信号対雑音比に応じて、スペクトルの偏倚に係る所定の閾値を変更する形態である。

図１９は、本発明の実施の形態５に係る非音声検出装置の一実施例である音声認識装置１について、制御手段２の音声認識処理の一例を示すフローチャートである。

【０１３１】

ステップＳ１３１乃至Ｓ１３５の処理は、夫々図３のステップＳ１１乃至Ｓ１５と同様であるので、説明を省略する。ステップＳ１３１乃至Ｓ１３５の処理で生成されてフレームバッファ４２に書き込まれたスペクトルの偏倚に対し、以下の処理が行われる。

【０１３２】

非音声区間検出部２２は、フレームバッファ４２を介してスペクトルの偏倚導出部２１より与えられたフレームについて、非音声区間を検出するサブルーチンを呼び出す（ステップＳ１３６）。その後、制御手段２は、非音声区間として検出されたフレームの音データ、及び非音声区間以外のフレームの音データに基づいて信号対雑音比を導出し（ステップＳ１３７）、導出した信号対雑音比の高／低に応じて、所定の閾値を下降／上昇させるように変更する（ステップＳ１３８）。

【０１３３】

音声区間判定部２３は、非音声区間検出部２２が非音声区間として検出できなかった区間を音声区間とみなし、そして、音声区間開始フレーム及び音声区間終了フレームを確定させて、一つの音声区間の判定を終える（ステップＳ１３９）。このようにして検出された音声区間は、フレームバッファを介して音声認識部２４に与えられる。

音声認識部２４は、音声認識の分野で一般的な技術を用いて、入力されたフレームバッファ４２の最後まで、音声認識処理を実行する（ステップＳ１４０）。

【０１３４】

その他、実施の形態１に対応する部分には同一符号を付して、それらの説明を省略する。

【０１３５】

このように、本発明の実施の形態５では、非音声区間として検出されたフレームの音データ、及び非音声区間以外のフレームの音データに基づいて信号対雑音比を導出し、導出した信号対雑音比の高／低に応じて、スペクトルの偏倚に係る所定の閾値を下降／上昇させるように変更する。

これにより、信号対雑音比が低下した場合に、雑音の影響により、スペクトルの偏倚が変動して、非音声区間を誤検出することを防止できる。

【０１３６】

実施の形態６

実施の形態６は、実施の形態１に対し、ピッチの各周波数成分の強度の最大値（以下、ピッチ強度という）を導出し、導出したピッチ強度に応じて、スペクトルの偏倚に係る所定の閾値を変更する形態である。

図２０及び図２１は、本発明の実施の形態６に係る非音声検出装置の一実施例である音声認識装置１について、非音声区間検出のサブルーチンに係る制御手段２の処理手順を示

すフローチャートである。

【０１３７】

非音声区間検出のサブルーチンが呼び出された場合、制御手段２は、そのときのポインタが示すフレームのピッチ強度を導出し（ステップＳ１５１）、導出したピッチ強度の大／小に応じて、所定の閾値を下降／上昇させるように変更する（ステップＳ１５２）。その後、制御手段２は、当該フレームのスペクトルの偏倚が、所定の閾値以上であるか否かを判定する（ステップＳ１５３）。所定の閾値未満であると判定した場合（ステップＳ１５３：ＮＯ）、制御手段２は、ワークメモリ４３に記憶されたフレームバッファ４２へのポインタを１フレーム後方に更新して（ステップＳ１５４）、リターンする。

これにより、制御手段２は、非音声区間を検出することなくリターンする。

【０１３８】

所定の閾値以上であると判定した場合（ステップＳ１５３：ＹＥＳ）、制御手段２は、そのときのポインタが示すフレームのフレーム番号を「開始フレーム番号」としてワークメモリ４３上に記憶する（ステップＳ１５５）。そして、制御手段２は、ワークメモリ４３上に設けた「フレームカウント」の記憶値を「１」に初期化する（ステップＳ１５６）。ここで、「フレームカウント」は、スペクトルの偏倚と所定の閾値との比較判定を行ったフレーム数を計数するものである。

【０１３９】

その後、制御手段２は、「フレームカウント」の記憶内容が所定数以上であるか否かを判定し（ステップＳ１５７）、所定数未満であると判定した場合（ステップＳ１５７：ＮＯ）、制御手段２は、「フレームカウント」の記憶内容に「１」を加算すると共に（ステップＳ１５８）、フレームバッファ４２へのポインタを１フレーム後方に更新する（ステップＳ１５９）。その後、制御手段２は、そのときのポインタが示すフレームのピッチ強度を導出し（ステップＳ１６０）、導出したピッチ強度に基づいて所定の閾値を変更する（ステップＳ１６１）。

【０１４０】

次いで、制御手段２は、スペクトルの偏倚が所定の閾値以上であるか否かを判定する（ステップＳ１６２）。所定の閾値以上であると判定した場合（ステップＳ１６２：ＹＥＳ）、制御手段２は、処理をステップＳ１５７に戻す。所定の閾値未満であると判定した場合（ステップＳ１６２：ＮＯ）、制御手段２は、「開始フレーム番号」の内容を消去して（ステップＳ１６３）、リターンする。

これにより、制御手段２は、非音声区間を検出することなくリターンする。

【０１４１】

ステップＳ１５７で「フレームカウント」の記憶内容が所定数以上と判定した場合（ステップＳ１５７：ＹＥＳ）、制御手段２は、非音声区間の終了フレームを検出する処理に移り、フレームバッファへのポインタを１フレーム後方に更新する（ステップＳ１６４）。その後、制御手段２は、そのときのポインタが示すフレームのピッチ強度を導出し（ステップＳ１６５）、導出したピッチ強度に基づいて所定の閾値を変更する（ステップＳ１６６）。

【０１４２】

次いで、制御手段２は、当該フレームのスペクトルの偏倚が所定の閾値以上であるか否かを判定する（ステップＳ１６７）。所定の閾値以上であると判定した場合（ステップＳ１６７：ＹＥＳ）、制御手段２は、処理をステップＳ１６４に戻す。所定の閾値未満であると判定した場合（ステップＳ１６７：ＮＯ）、制御手段２は、そのときのポインタが示すフレームの１つ前のフレーム番号を「終了フレーム番号」としてワークメモリ４３上に記憶し（ステップＳ１６８）、リターンする。

これにより、「開始フレーム番号」及び「終了フレーム番号」で区切られた区間が、検出された非音声区間となる。

【０１４３】

ここで、図２０図２１を用いて説明したステップＳ１５１、Ｓ１６０及びＳ１６５にお

けるピッチ強度について詳述する。
ピッチ強度Ｂは、短時間スペクトルS(ω)の自己相関関数γ(τ)を用いて、以下の式９を用いて導出することができる。
【０１４４】
Ｂ＝argmaxγ(τ)，１≦τ≦τmax、　　・・・・・・式９
但し、τmaxは、想定される最高ピッチ周波数に対応する値。
【０１４５】
例えば、８０００Hzサンプリングで、１フレーム長が２５６サンプルの場合、短時間スペクトルは、０～４０００Hzを１２９次元ベクトルで表現できる。この場合、最高ピッチ周波数を５００Hzとしたとき、短時間スペクトル上では、５００／４０００×１２８＝１６より、τmax＝１６となる。
【０１４６】
その他、実施の形態１に対応する部分には同一符号を付して、それらの説明を省略する。
【０１４７】
このように、本発明の実施の形態６では、各フレームの音データについて、ピッチ強度を導出し、導出したピッチ強度の大／小に応じて、スペクトルの偏倚に係る所定の閾値を下降／上昇させる。例えば、ピッチ強度が大きい場合、即ち、ピッチが明確に現れている場合は、音データが音声の母音又は半母音であることが想定される。この場合、スペクトルの偏倚が取り得る値は制限される。従って所定の閾値を下げて非音声区間を検出する判定条件を緩めても、誤検出を抑止して高精度に非音声区間を検出することができる。
【０１４８】
尚、導出したピッチ強度に応じて所定の閾値を変更するのではなく、例えば下記（ｈ）の判定を加えてもよい。
（ｈ）：ピッチ強度Ｂ≧所定の強度、且つ、|Ａ|≧０．５が０．５秒以上継続する場合、当該区間は非音声とする。（上述した（ｂ）又は（ｃ）の判定とピッチ強度とを組合せて改良したもの）
【０１４９】
実施の形態７
実施の形態７は、実施の形態１において、スペクトルの偏倚に係る所定の閾値を、事前の学習によって決定する形態である。
図２２は、本発明の実施の形態７に係る非音声検出装置の一実施例である音声認識装置１について、制御手段２の音声認識処理の一例を示すフローチャートである。
【０１５０】
ステップＳ１７１乃至Ｓ１７４の処理は、夫々図３のステップＳ１１乃至Ｓ１４と同様であるので、説明を省略する。ステップＳ１７１乃至Ｓ１７４の処理で生成された各フレームに対し、以下の処理が行われる。
【０１５１】
制御手段２は、フレームバッファ４２を介して与えられたフレームについて、音データにおける発声区間をマーキングする（ステップＳ１７５）。この場合、学習用の音声データには、音素ラベリングがされているため、容易に発声区間をマーキングすることが可能である。更に、制御手段２は、スペクトルの偏倚|Ａ|が取り得る値の範囲〔－１,－１〕内にＮ個の閾値を設定する（ステップＳ１７６）。そして、制御手段２は、Ｎ個の閾値のうち１つの閾値について、当該閾値以上となるフレームが継続する最大数を集計する（ステップＳ１７７）。
【０１５２】
次いで、制御手段２は、Ｎ個の閾値全てについての集計を終了したか否かを判定する（ステップＳ１７８）。未集計の閾値があると判定した場合（ステップＳ１７８：ＮＯ）、制御手段２は、処理をステップＳ１７７に戻す。Ｎ個の閾値全てについての集計を終了したと判定した場合（ステップＳ１７８：ＹＥＳ）、制御手段２は、集計した結果に基づい

て、スペクトルの偏倚に係る所定の閾値を決定する（ステップＳ１７９）。

この場合、所定の閾値を大きめに（又は小さめに）決定して、非音声区間の誤検出を抑止することが好ましい。

【０１５３】

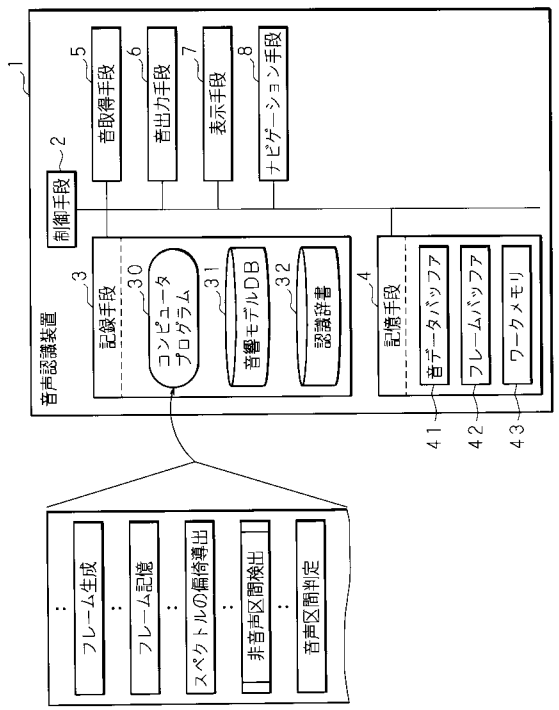
このように、本発明の実施の形態７では、既存の音声データのマーキングされた発声区間について、予め複数の閾値候補を準備し、所定の閾値以上となるフレームが継続する最大数を集計した結果に基づいて、複数の閾値候補の中から、スペクトルの偏倚に係る所定の閾値の最適値を決定する。

これにより、高精度に非音声区間を検出することができる。

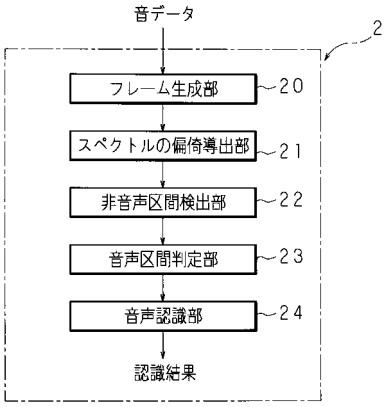
【０１５４】

実施の形態１乃至７にあっては、高域・低域強度の絶対値|Ａ|をスペクトルの偏倚とし、スペクトルの偏倚が所定の正の閾値以上であるか否かを判定する場合について説明したが、高域・低域強度Ａをスペクトルの偏倚とし、スペクトルの偏倚が正の値（又は負の値）の場合、所定の正の閾値以上（又は所定の負の閾値以下）であるか否かを判定するようにしてもよい。

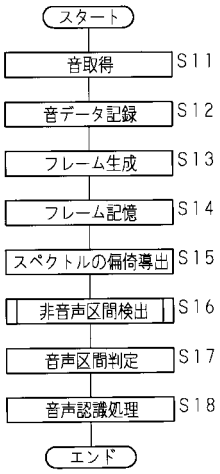
【図１】



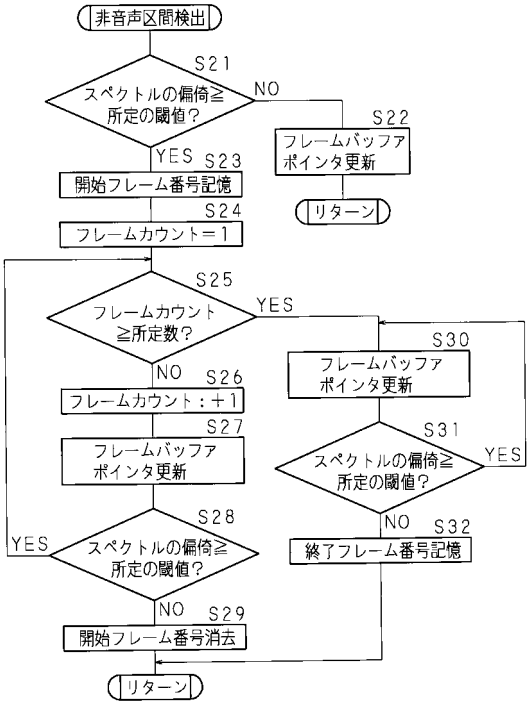
【図２】



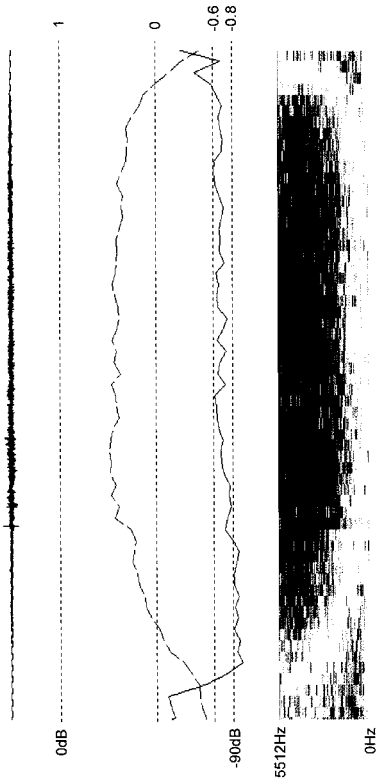
【図３】


スタート
音取得
S11
音データ記録
S12
フレーム生成
S13
フレーム記憶
S14
スペクトルの偏倚導出
S15
非音声区間検出
S16
音声区間判定
S17
音声認識処理
S18
エンド

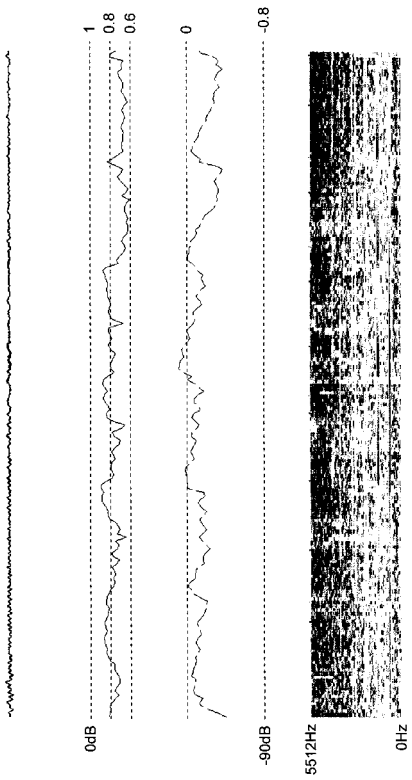
【図４】


非音声区間検出
S21
スペクトルの偏倚≧所定の閾値？
NO
S22
フレームバッファポインタ更新
リターン
YES
S23
開始フレーム番号記憶
S24
フレームカウント＝1
S25
フレームカウント≧所定数？
YES
S30
フレームバッファポインタ更新
NO
S26
フレームカウント：＋1
S27
フレームバッファポインタ更新
S31
スペクトルの偏倚≧所定の閾値？
YES
NO
S32
終了フレーム番号記憶
S28
YES
スペクトルの偏倚≧所定の閾値？
NO
S29
開始フレーム番号消去
リターン

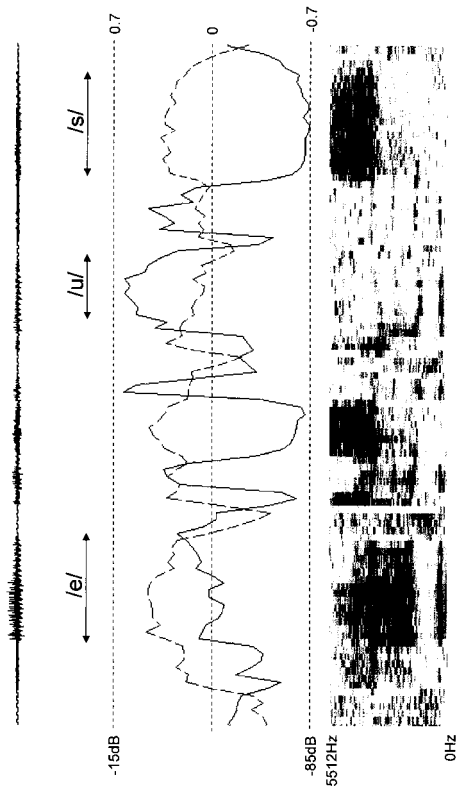
【図５】


1
0
-0.6
-0.8
0dB
-90dB
5512Hz
0Hz

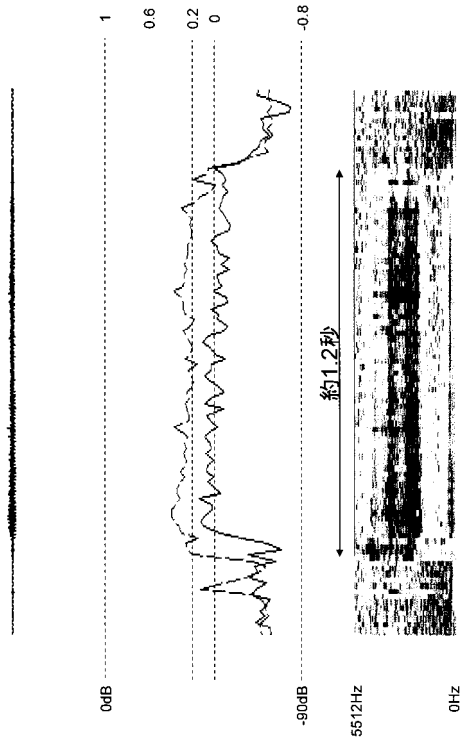
【図６】


1
0.8
0.6
0
-0.8
0dB
-90dB
5512Hz
0Hz

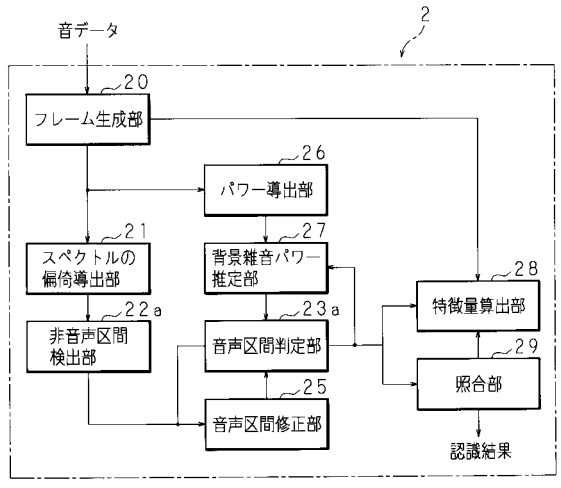
【図７】


/s/
/u/
/e/
0.7
0
-0.7
-15dB
-85dB
5512Hz
0Hz

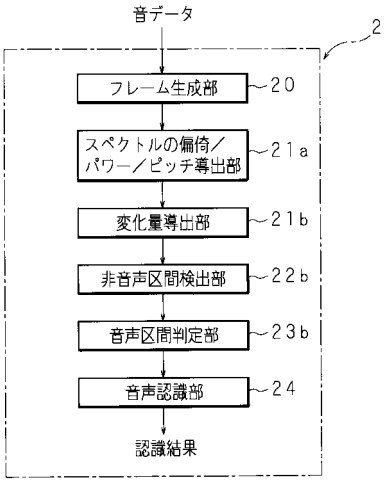
【図８】


1
0.6
0.2
0
-0.8
約1.2秒
0dB
-90dB
5512Hz
0Hz

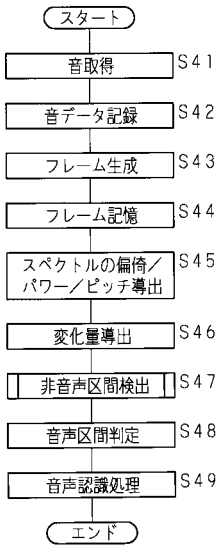
【図９】


音データ
2
20
フレーム生成部
26
パワー導出部
21
スペクトルの偏倚導出部
27
背景雑音パワー推定部
28
特徴量算出部
22a
非音声区間検出部
23a
音声区間判定部
29
照合部
25
音声区間修正部
認識結果

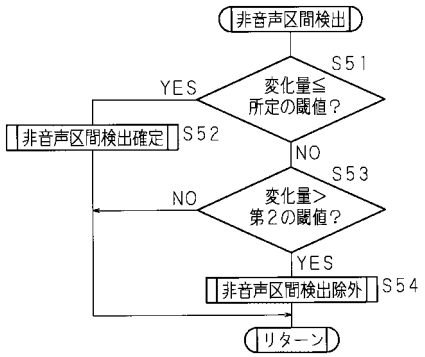
【図１０】


音データ
2
フレーム生成部
20
スペクトルの偏倚／パワー／ピッチ導出部
21a
変化量導出部
21b
非音声区間検出部
22b
音声区間判定部
23b
音声認識部
24
認識結果

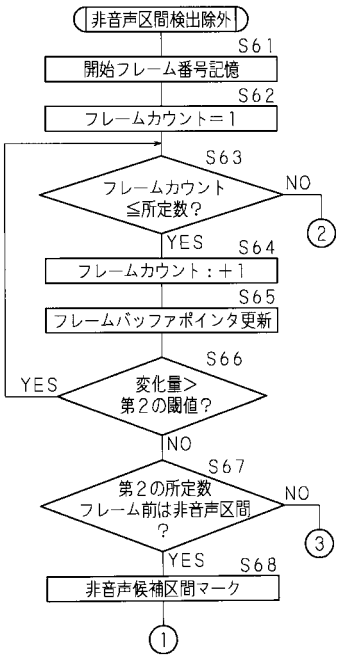
【図１１】



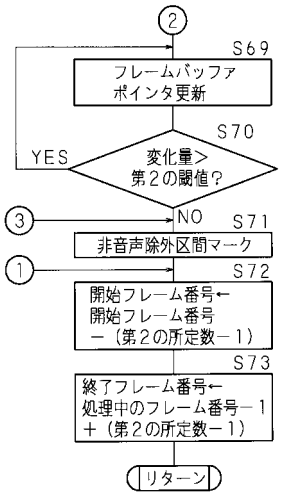
【図１２】



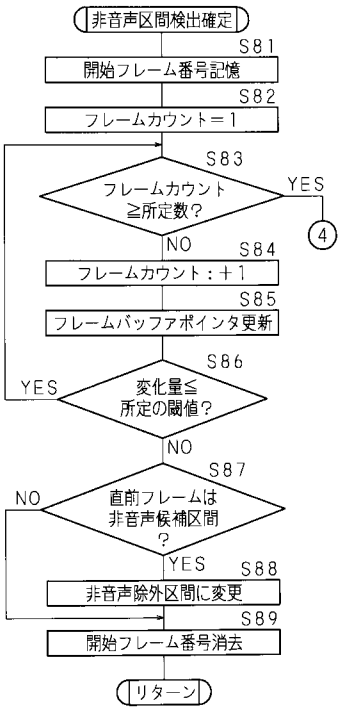
【図１３】



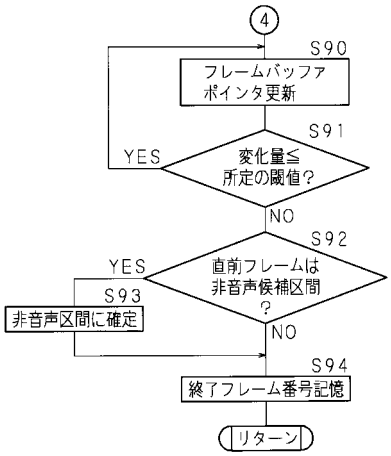
【図１４】



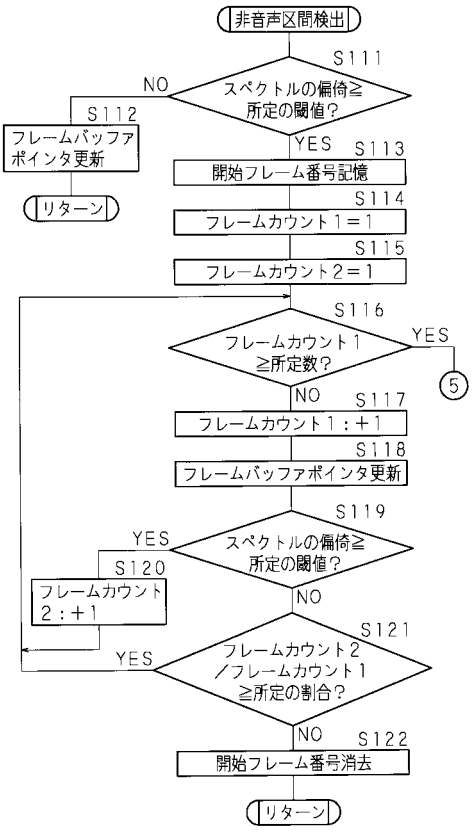
【図１５】



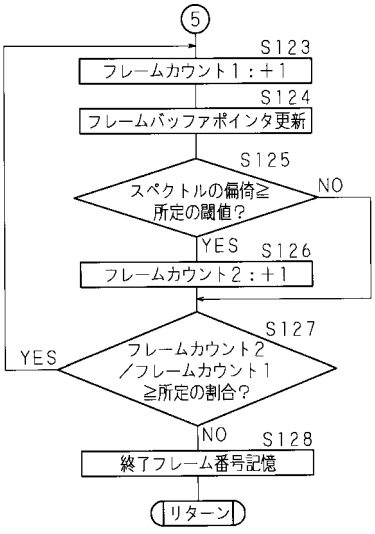
【図１６】



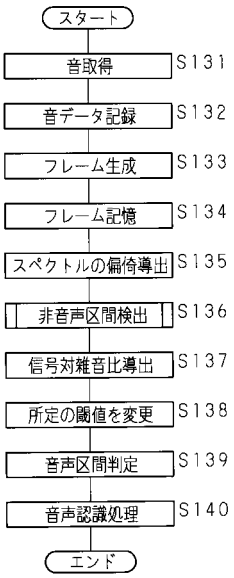
【図１７】



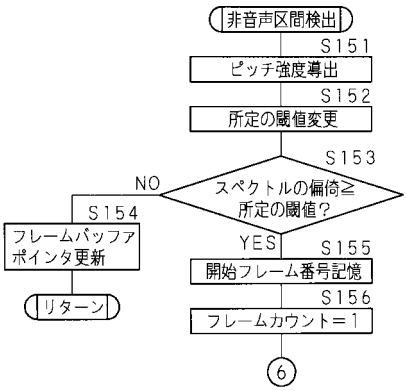
【図１８】



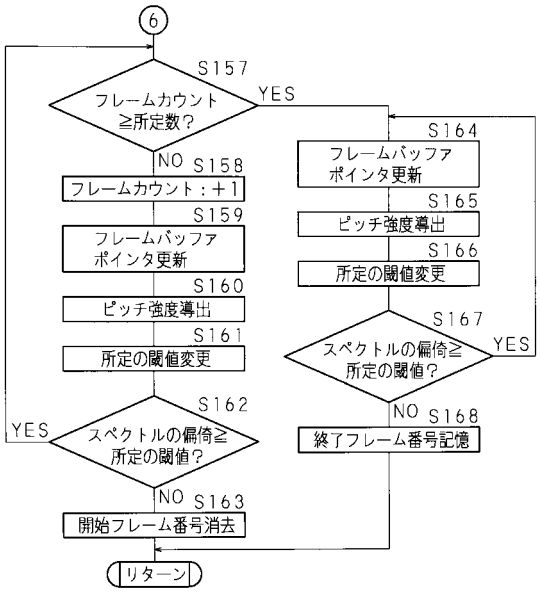
【図１９】



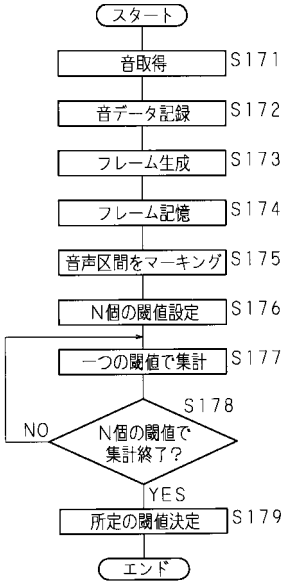
【図２０】



【図２１】



【図２２】



フロントページの続き

(56)参考文献 特開平０９－１５２８９４（ＪＰ，Ａ）
特開２００１－３５０４８８（ＪＰ，Ａ）
特開平０６－０８３３９１（ＪＰ，Ａ）
特開２００５－１５６８８７（ＪＰ，Ａ）
特開２００６－２０９０６９（ＪＰ，Ａ）
特開２００７－２３３２６７（ＪＰ，Ａ）
特開平０７－２１２２９６（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ１０Ｌ　１５／０４，２５／７８
ＪＳＴＰｌｕｓ（ＪＤｒｅａｍＩＩ）