

(19) 世界知的所有権機関
国際事務局



(43) 国際公開日
2009年6月25日 (25.06.2009)

PCT

(10) 国際公開番号
WO 2009/078093 A1

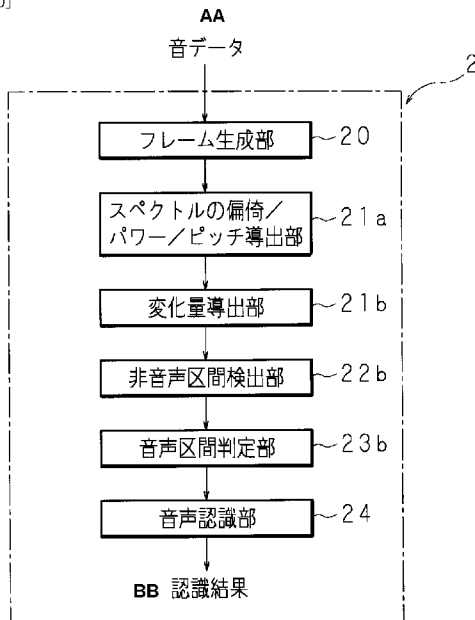
- (51) 国際特許分類:
G10L 11/00 (2006.01)
- (21) 国際出願番号: PCT/JP2007/074274
- (22) 国際出願日: 2007年12月18日 (18.12.2007)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (71) 出願人 (米国を除く全ての指定国について): 富士通株式会社 (FUJITSU LIMITED) [JP/JP]; 〒2118588 神奈川県川崎市中原区上小田中4丁目1番1号 Kanagawa (JP).
- (72) 発明者; および
- (75) 発明者/出願人 (米国についてのみ): 鷲尾 信之
- (74) 代理人: 河野 登夫 (KOHNO, Takao); 〒5400035 大阪府大阪市中央区釣鐘町二丁目4番3号 河野特許事務所 Osaka (JP).
- (81) 指定国 (表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

[続葉有]

(54) Title: NON-SPEECH SECTION DETECTING METHOD AND NON-SPEECH SECTION DETECTING DEVICE

(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

[図10]



AA SOUND DATA
 20 FRAME GENERATING UNIT
 21a SPECTRUM BIAS/POWER/PITCH DERIVING UNIT
 21b VARIATION AMOUNT DERIVING UNIT
 22b NON-SPEECH SECTION DETECTING UNIT
 23b SPEECH SECTION JUDGING UNIT
 24 SPEECH RECOGNIZING UNIT
 BB RECOGNITION RESULT

(57) Abstract: A frame generating section (20) of control means (2) generates a frame having a predetermined time length from sound data. A spectrum bias/power/pitch deriving unit (21a) derives at least one of the bias of the spectrum obtained by converting the sound data into a component on the frequency axis, the power of the sound data, and the pitch of the sound data. A variation amount deriving unit (21b) derives the amount of variation of the value derived by the spectrum bias/power/pitch deriving unit (21a) from that of the previous frame. As the bias of the spectrum, the ratio of the first-order autocorrelation function of the sound data to the zero-order autocorrelation function thereof is used. If the amount of variation is judged to be a predetermined threshold or less, a non-speech section detecting unit (22b) detects a non-speech section including consecutive frames when the amount of variation is judged to be a predetermined threshold or less and when the number of consecutive frames is a predetermined one or more. The section where the amount of variation is large singly is excluded from the non-speech section. If the section where the amount of variation is large singly is sandwiched between two non-speech sections, the section is detected as a non-speech section irrespective of the judgment.

(57) 要約: 制御手段2のフレーム生成部20は、音データから所定の時間長のフレームを生成し、スペクトルの偏倚/パワー/ピッチ導出部21aは、音データを周波数軸上の成分に変換し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を導出し、更に、導出された値について、変化量導出部21bが、前フレームとの変化量を導出する。スペクトルの偏倚としては、音データの0次の自己相関関

数に対する1次の自己相関関数の比を用いる。非音声区間検出部22bは、前記変化量が所定の閾値以下であると判定した場合、該判定の成立するフレームが連なる数が所定数以上であるときに、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。ま

[続葉有]

WO 2009/078093 A1



(84) 指定国 (表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, PL, PT, RO, SE, SI, SK,

TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

添付公開書類:
— 国際調査報告書

明 細 書

非音声区間検出方法及び非音声区間検出装置

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の偏倚を導出し、導出した偏倚が、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)であるか否かを判定し、前記閾値以上(又は前記閾値以下)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の偏倚、並びに各フレームの音データのパワー及びピッチの少なくとも一を指標として導出し、導出した指標について、前フレームとの変化量を導出し、導出した変化量が、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の偏倚を導出する導出手段と、導出した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)であるか否かを判定する判定手段と、前記閾値以上(又は前記閾値以下)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の偏倚、並びに各フレームの音データのパワー及びピッチの少なくとも一を指標として導出する導出手段と、導出した指標について、前フレームとの変化量を導出する第2の導出手段と、導出した変化量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] 第7の非音声区間検出装置は、第4の装置乃至第6の装置の何れかにおいて、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8の非音声区間検出装置は、第3の装置乃至第7の装置の何れかにおいて、前記スペクトルの偏倚は、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] 第9の非音声区間検出装置は、第3の装置乃至第8の装置の何れかにおいて、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] 第10の非音声区間検出装置は、第3の装置乃至第9の装置の何れかにおいて、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] 第11の非音声区間検出装置は、第3の装置乃至第10の装置の何れかにおいて、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] 第12の非音声区間検出装置は、第3の装置乃至第11の装置の何れかにおいて、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] 第13の非音声区間検出装置は、第3の装置乃至第12の装置の何れかにおいて、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] 第14の非音声区間検出装置は、第3の装置乃至第13の装置の何れかにおいて、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] 第15の非音声区間検出装置は、第3の装置乃至第13の装置の何れかにおいて、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] 第7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8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、これをスペクトルの偏倚とすることにより、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] 第9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变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] 第10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] 第11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] 第12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] 第13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] 第14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] 第15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)であるかを判定し、前記閾値以上(又は前記閾値以下)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。

[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、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偏倚、並びに各フレームの音データの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、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]

$$c(\tau) = \frac{1}{N-1} \sum_{n=0}^{N-2} x(n)x(n+\tau) \quad , \quad 0 \leq \tau \leq 1 \quad \dots \text{式 1}$$

[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]

$$c(\tau) = \sum_{n=0}^{N-2} x(n)x(n+\tau) \quad , \quad 0 \leq \tau \leq 1 \quad \dots \text{式 2}$$

[0058] また、自己相関関数 $c(\tau)$ は、Wiener-Khintchineの定理により、短時間スペクトル $S(\omega)$ 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 $S(\omega)$ 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 $c(\tau)$ 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。

$$[0060] \quad A = c(1)/c(0) \quad (c(0) \neq 0) \quad \dots \text{式 3}$$

$$A = 0 \quad (c(0) = 0) \quad \dots \text{式 4}$$

[0061] この場合、 A は、 $-1 \leq A \leq 1$ 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ます、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、 $|A| \geq 0.7$ 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 $|A| \geq 0.5$ 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) : $|A| \geq 0.7$ が 0.1 秒以上継続する場合、当該区間は非音声とする。

(b) : $|A| \geq 0.5$ 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) : $|A| \geq 0.6$ 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、標準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] $C(t) = |A(t) - A(t-1)|, t > 1$ 式5

$C(t) = 0, t = 1$ 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): $C(t) \leq 0.05$ 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): $C(t) \leq 0.1$ 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、変化量が多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、エアコンの風量変動して、定量的な雑音に変化した瞬間)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にあつては、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つてもよい。

[0115] [数3]

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

[数4]

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

[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)にあつては、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つた場合に、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] 次に、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 $S(\omega)$ の自己相関関数 $\gamma(\tau)$ を用いて、以下の式9を用いて導出することができる。

[0144] $B = \operatorname{argmax} \gamma(\tau), 1 \leq \tau \leq \tau_{\max}, \dots\dots\dots$ 式9

但し、 τ_{\max} 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、 $500/4000 \times 128 = 16$ より、 $\tau_{\max} = 16$ 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 $B \geq$ 所定の強度、且つ、 $|A| \geq 0.5$ 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にあつては、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。

請求の範囲

- [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の偏倚を導出し、導出した偏倚が、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)であるか否かを判定し、
- 前記閾値以上(又は前記閾値以下)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。
- [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の偏倚、並びに各フレームの音データのパワー及びピッチの少なくとも一を指標として導出し、
- 導出した指標について、前フレームとの変化量を導出し、
- 導出した変化量が、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。
- [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の偏倚を導出する導出手段と、

導出した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)であるか否かを判定する判定手段と、

前記閾値以上(又は前記閾値以下)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。

- [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の偏倚、並びに各フレームの音データのパワー及びピッチの少なくとも一を指標として導出する導出手段と、

導出した指標について、前フレームとの変化量を導出する第2の導出手段と、

導出した変化量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。

- [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

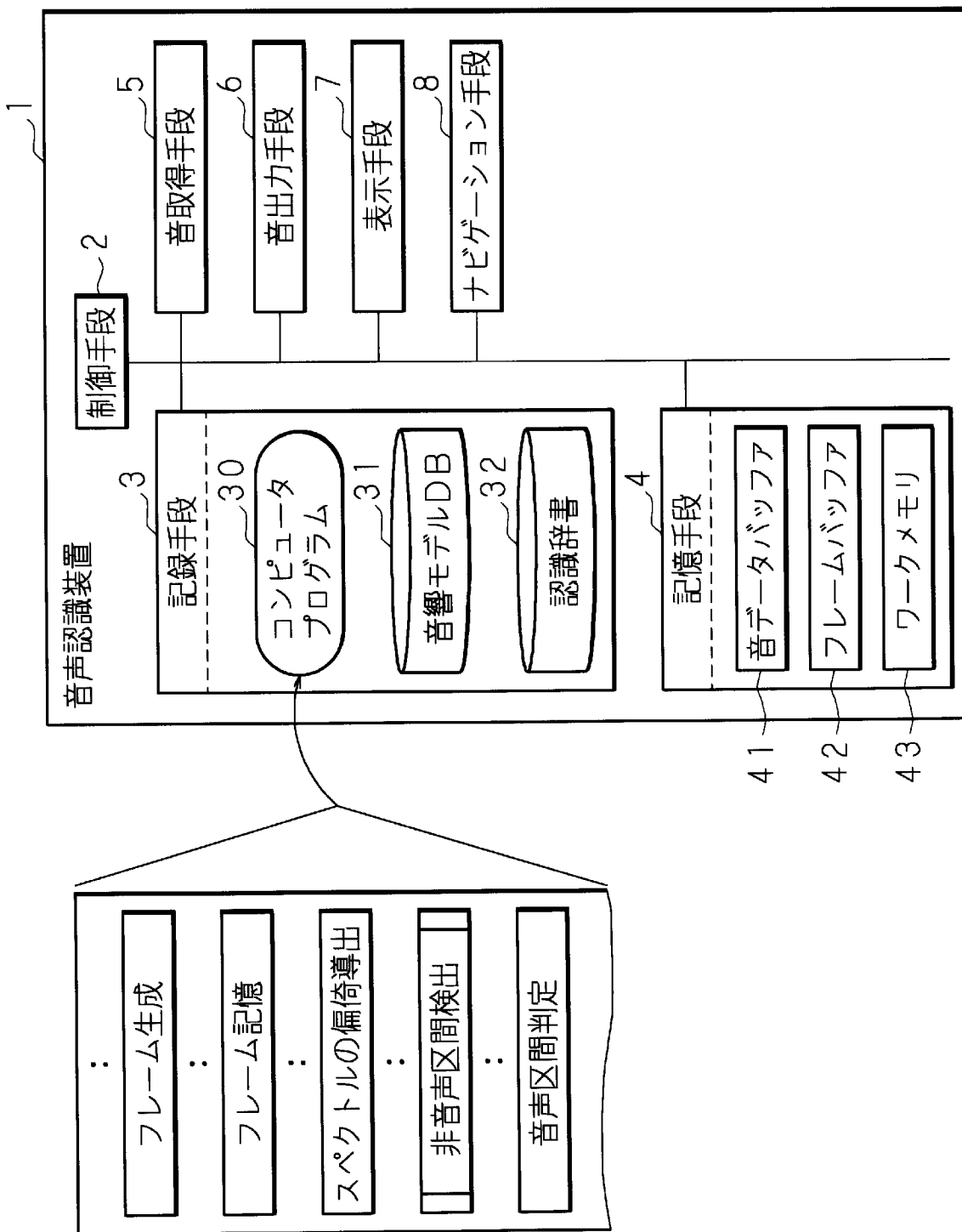
記載の非音声区間検出装置。

- [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。
- [7] 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特徴とする請求項4乃至6の何れかに記載の非音声区間検出装置。
- [8] 前記スペクトルの偏倚は、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7の何れかに記載の非音声区間検出装置。
- [9] 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特徴とする請求項3乃至8の何れかに記載の非音声区間検出装置。
- [10] 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
を備えることを特徴とする請求項3乃至9の何れかに記載の非音声区間検出装置。

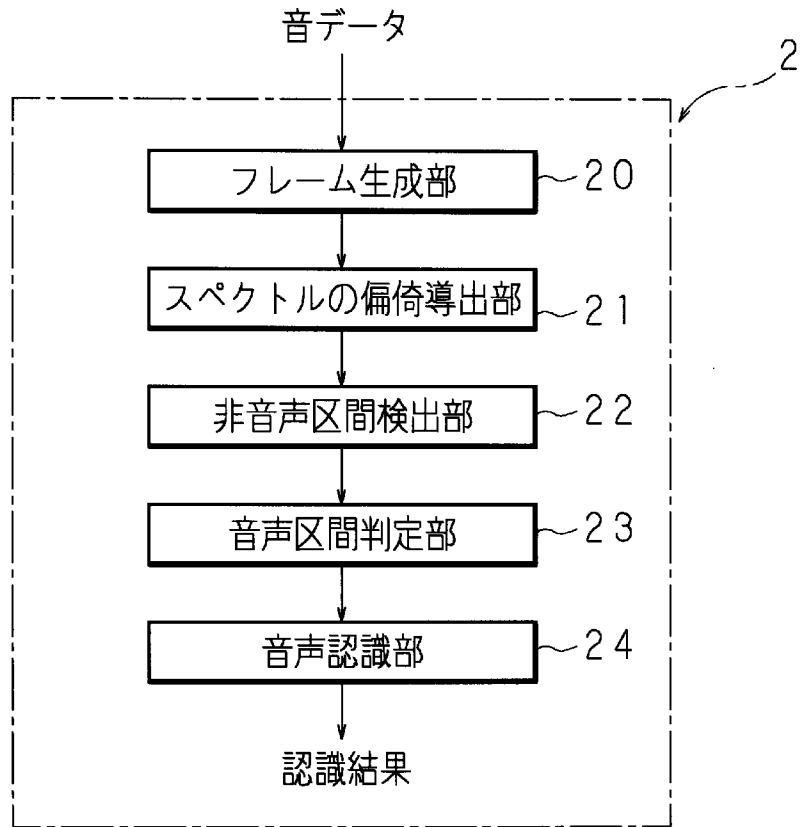
- [11] 非音声区間として検出されたフレームの音データ、及び前記非音声区間以外のフレームの音データに基づいて、信号対雑音比を導出する手段と、
導出した信号対雑音比に基づいて、前記閾値を変更する手段と
を備えることを特徴とする請求項3乃至10の何れかに記載の非音声区間検出装置
。
- [12] 各フレームの音データについて、ピッチの各周波数成分の強度の最大値を導出する手段と、
導出した強度の最大値に基づいて、前記閾値を変更する手段と
を備えることを特徴とする請求項3乃至11の何れかに記載の非音声区間検出装置
。
- [13] 人が発声した音データについて、予め準備された複数の候補閾値に対し、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
を備えることを特徴とする請求項3乃至12の何れかに記載の非音声区間検出装置
。
- [14] 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特徴とする請求項3乃至13の何れかに記載の非音声区間検出装置。

- [15] 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特徴とする請求項3乃至13の何れかに記載の非音声区間検出装置。

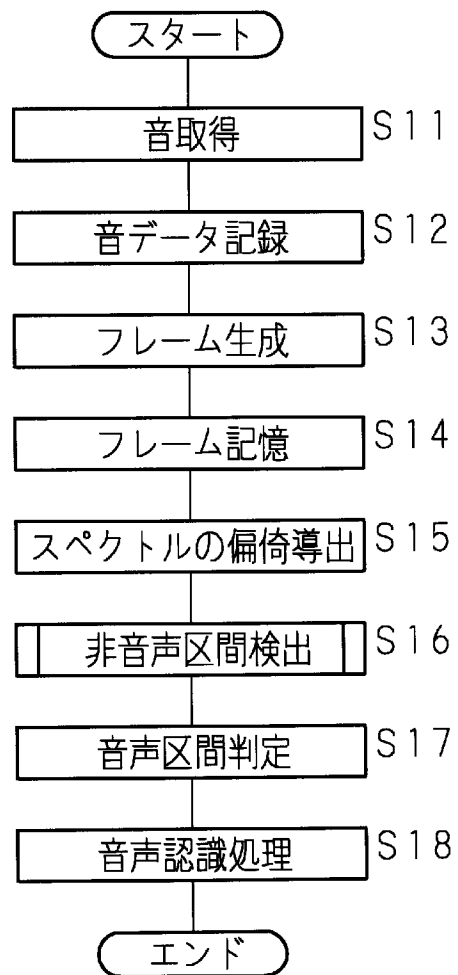
[図1]



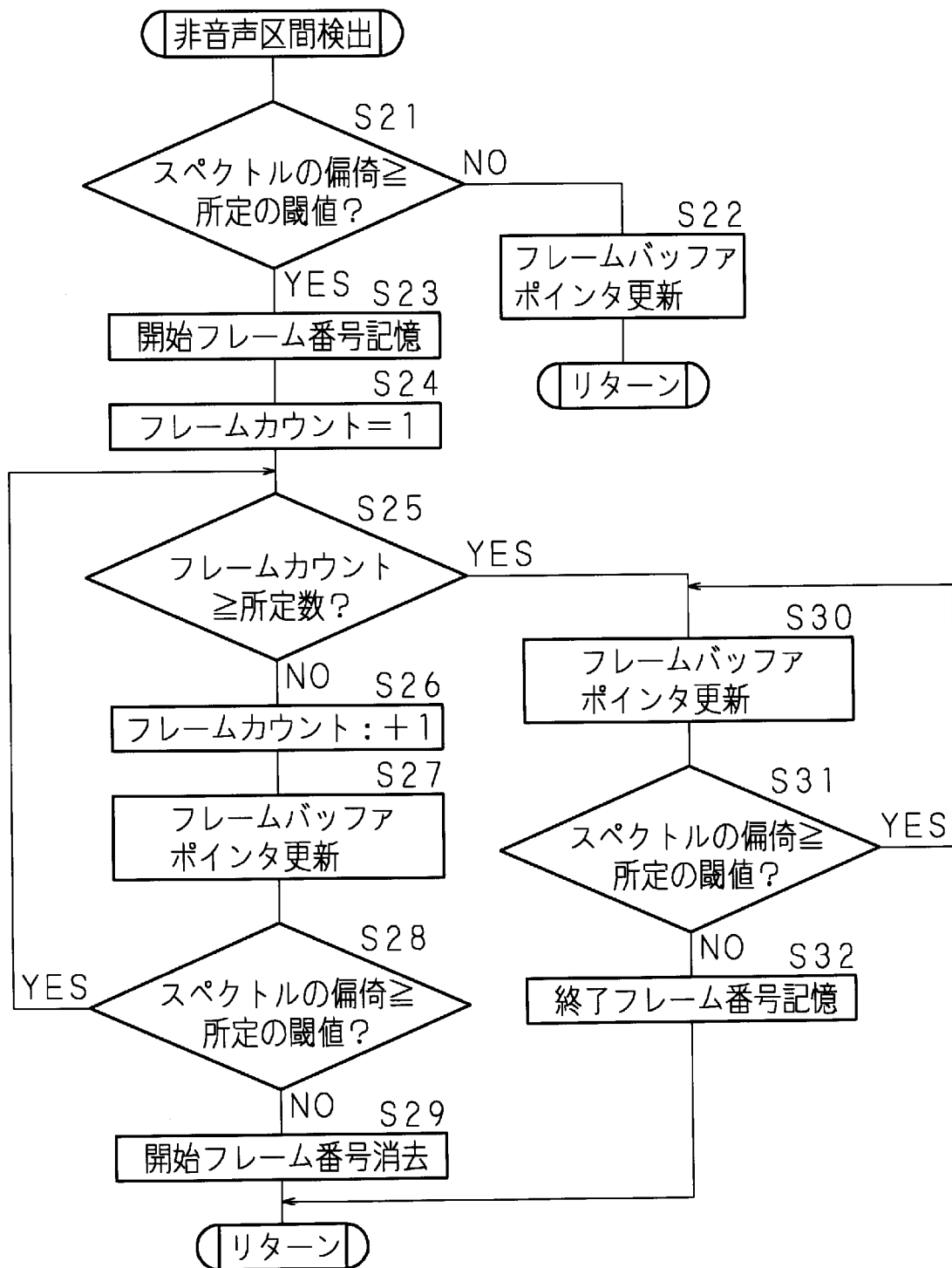
[図2]



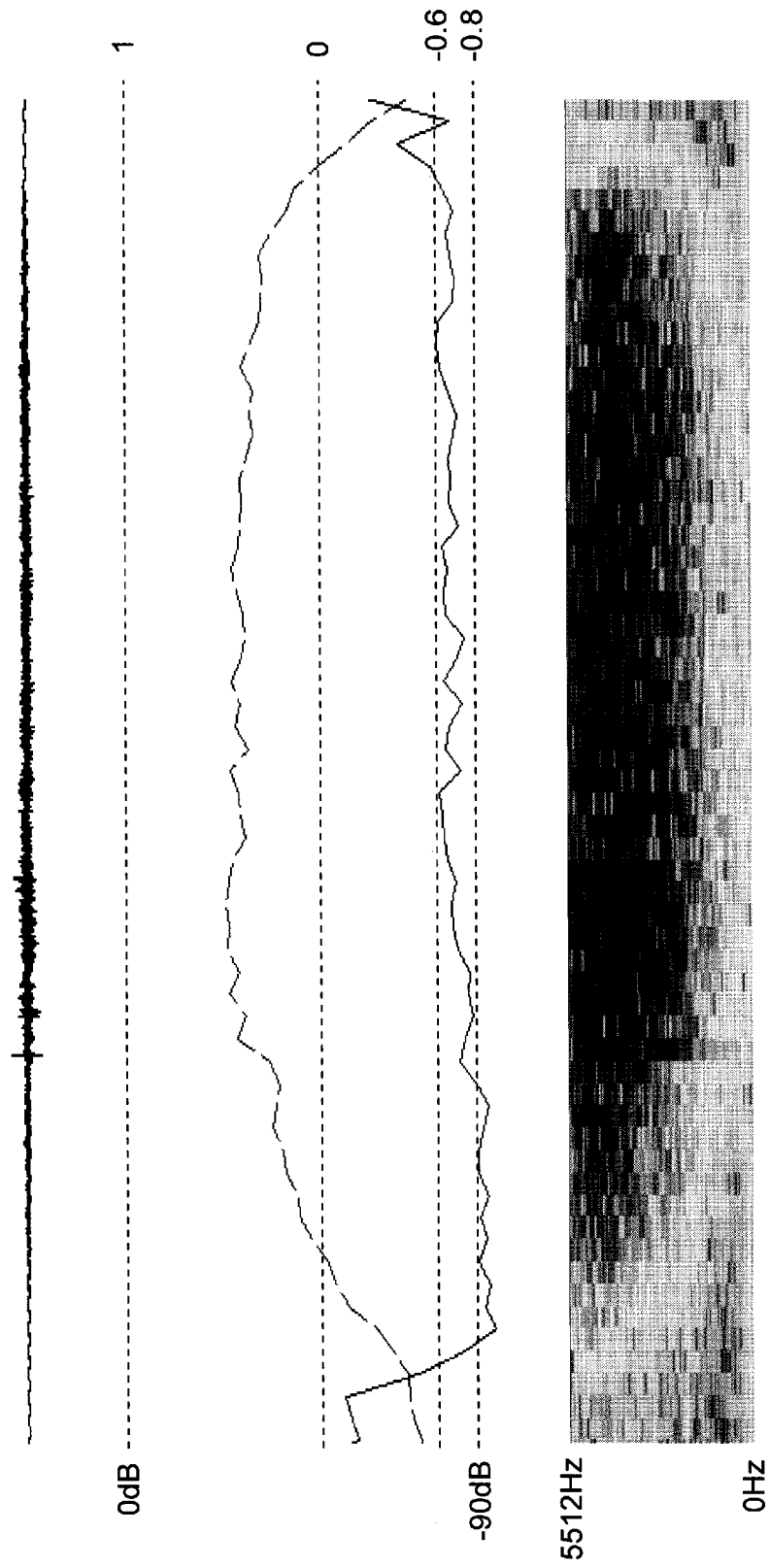
[図3]



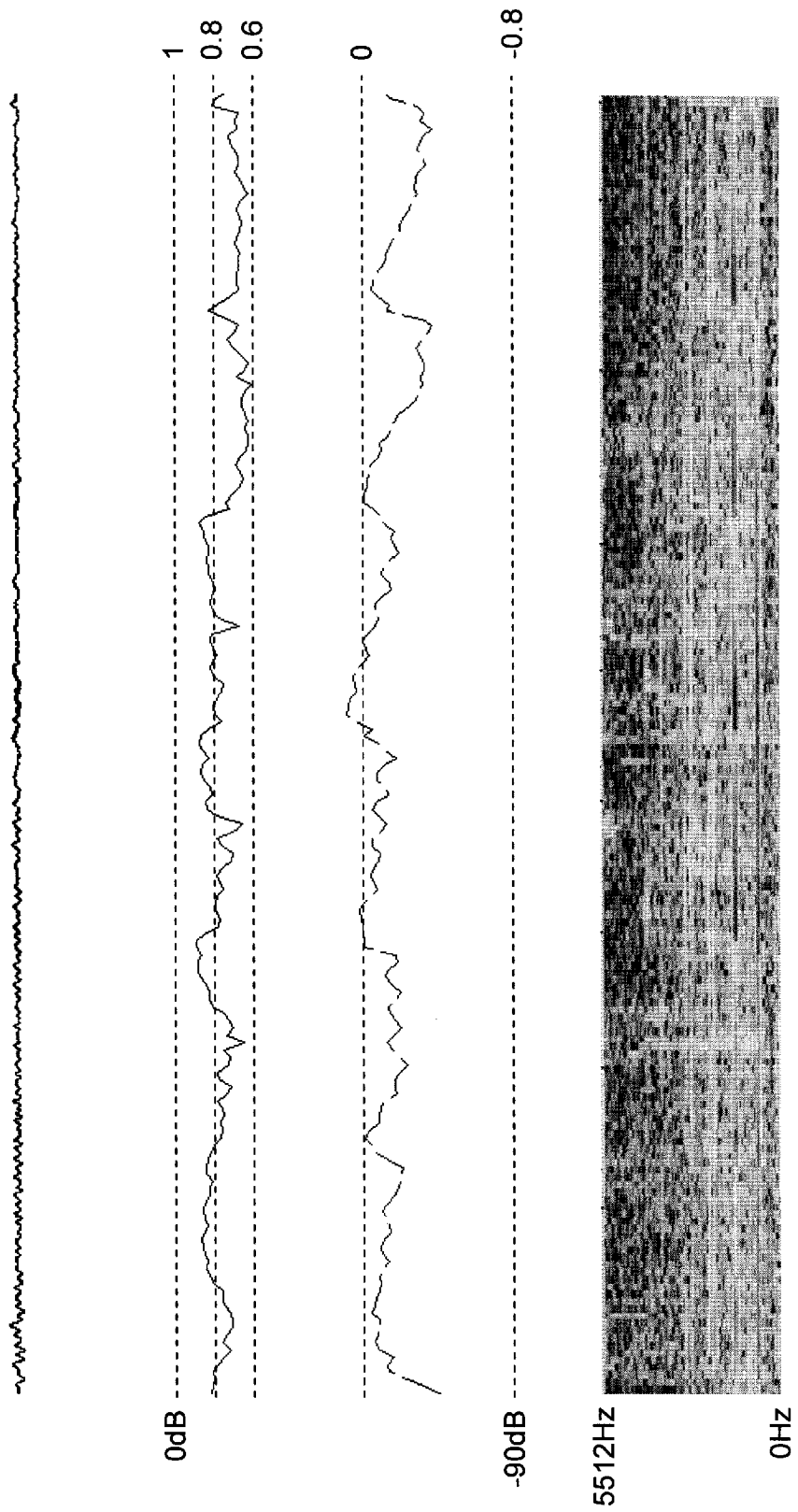
[図4]



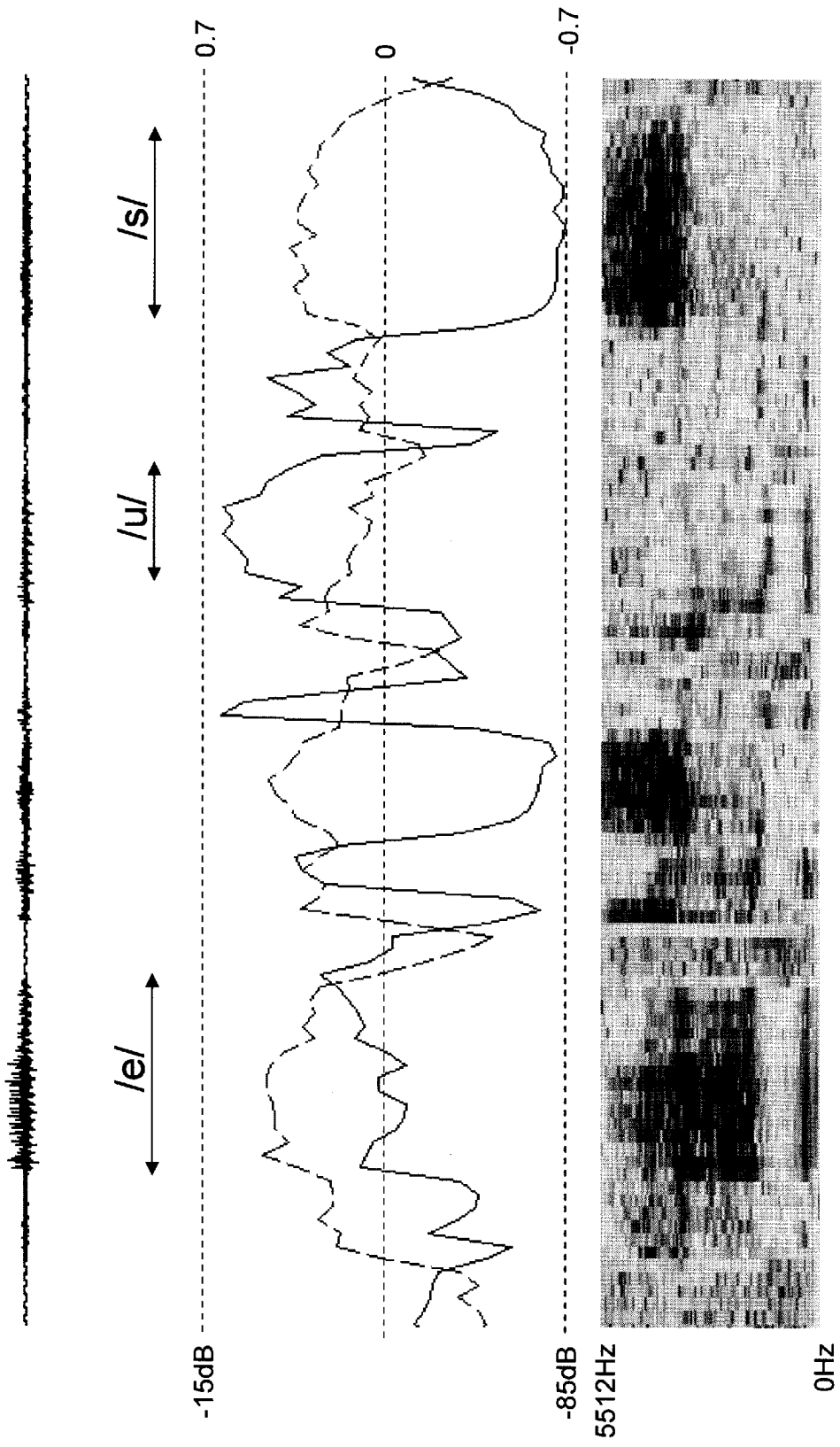
[図5]



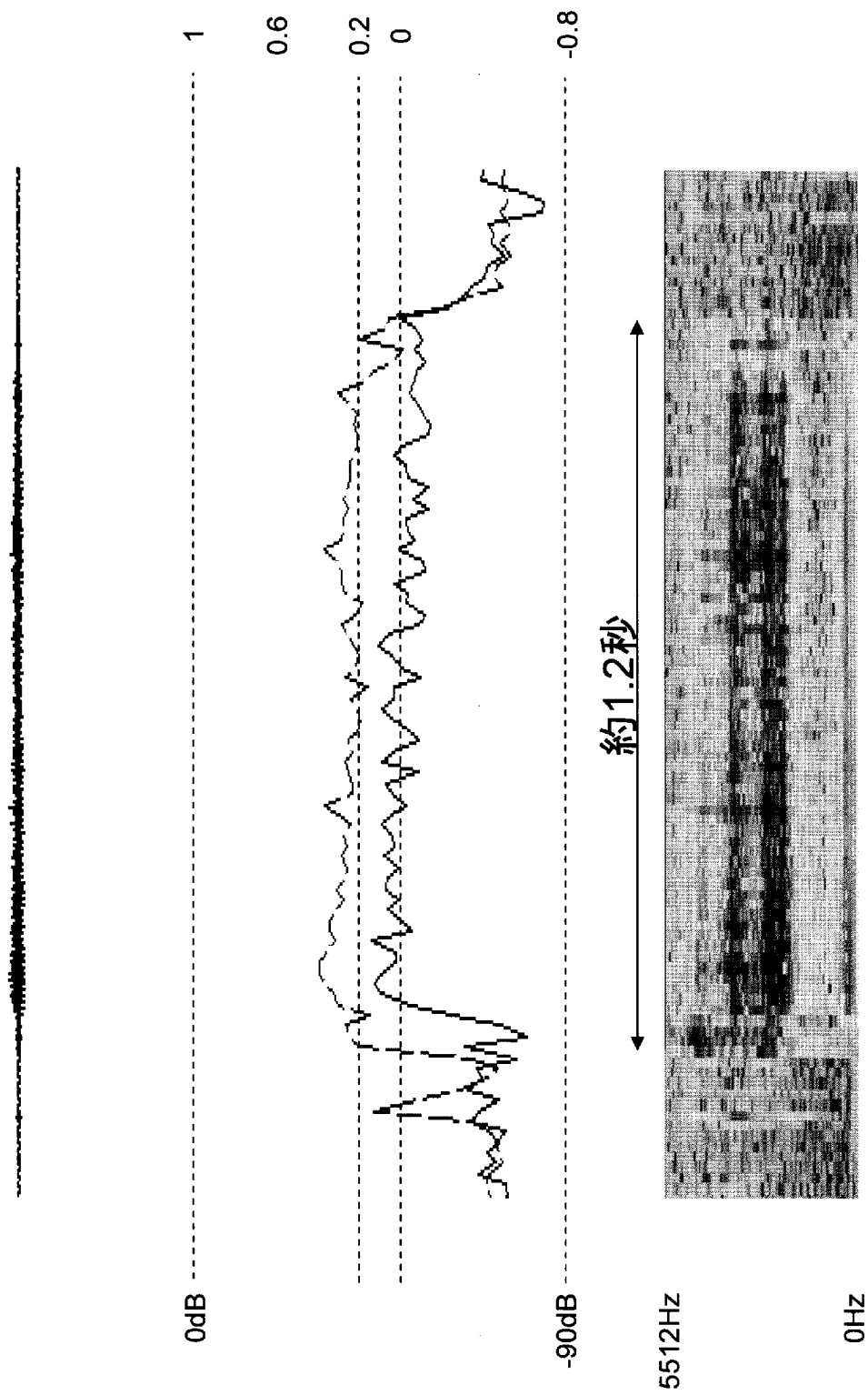
[図6]



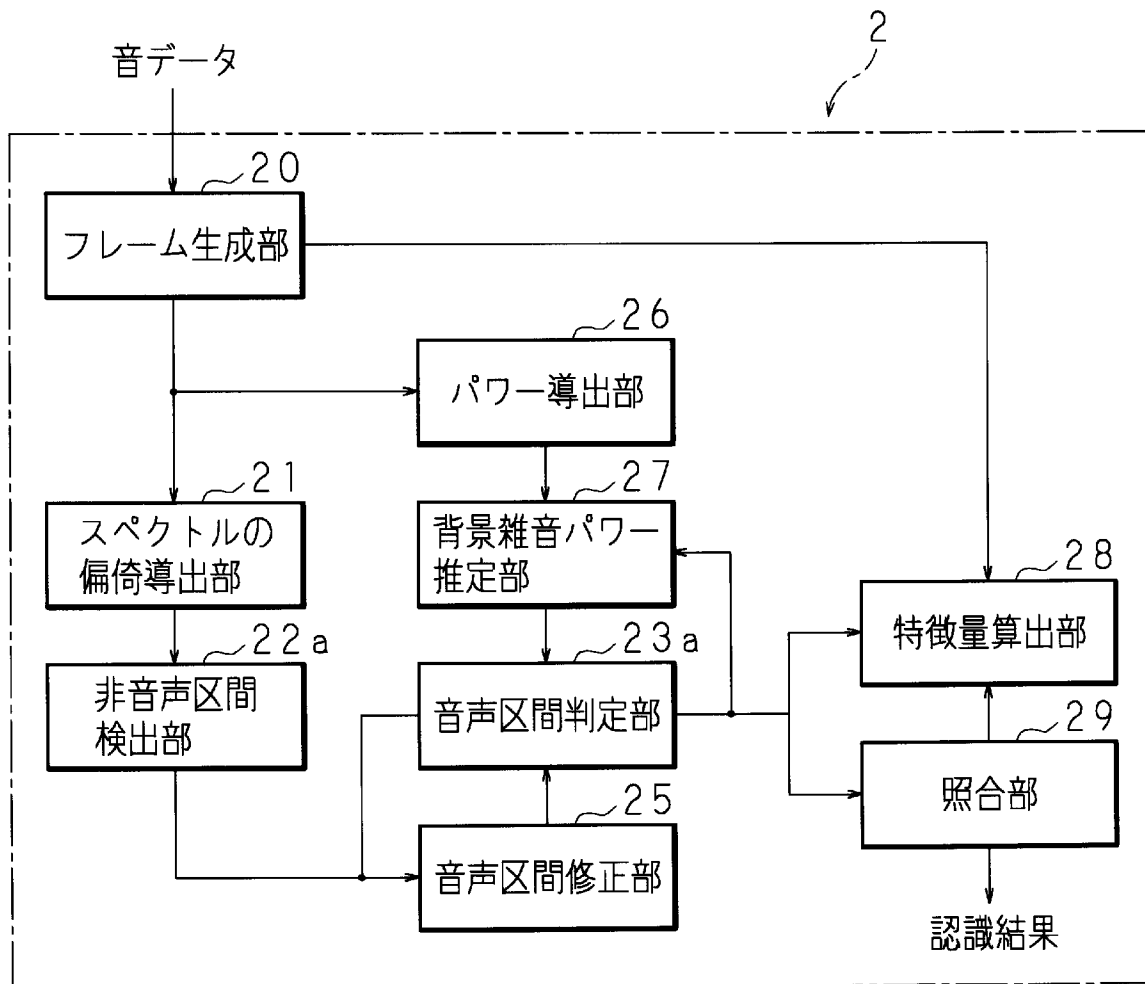
[図7]



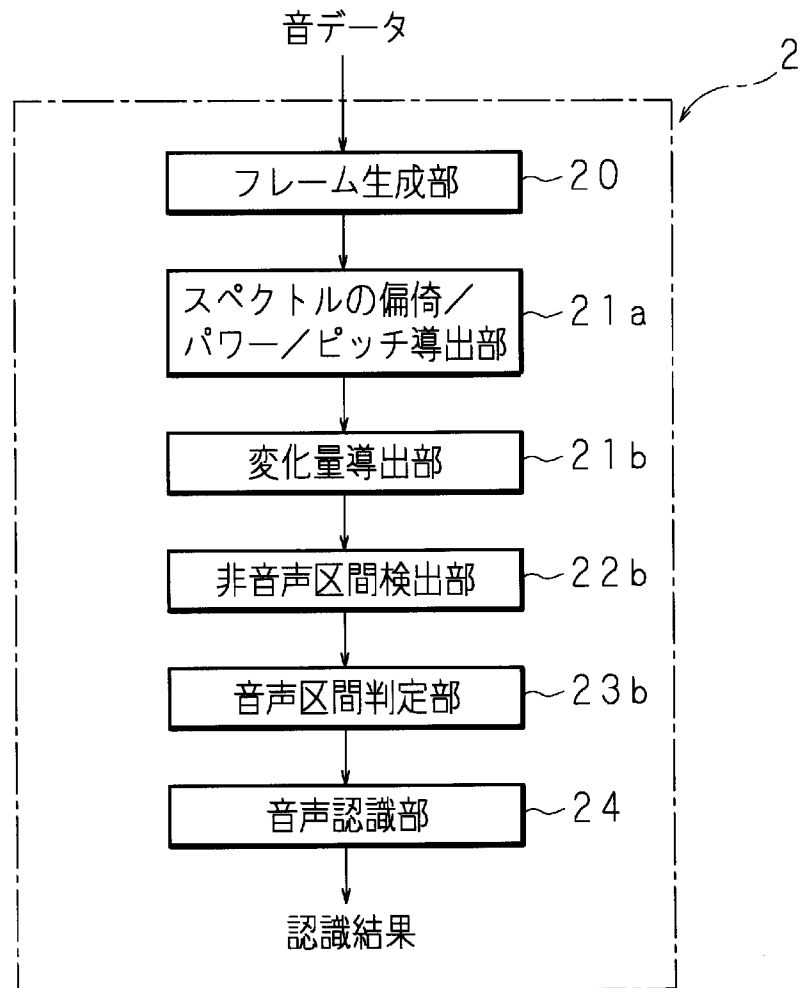
[図8]



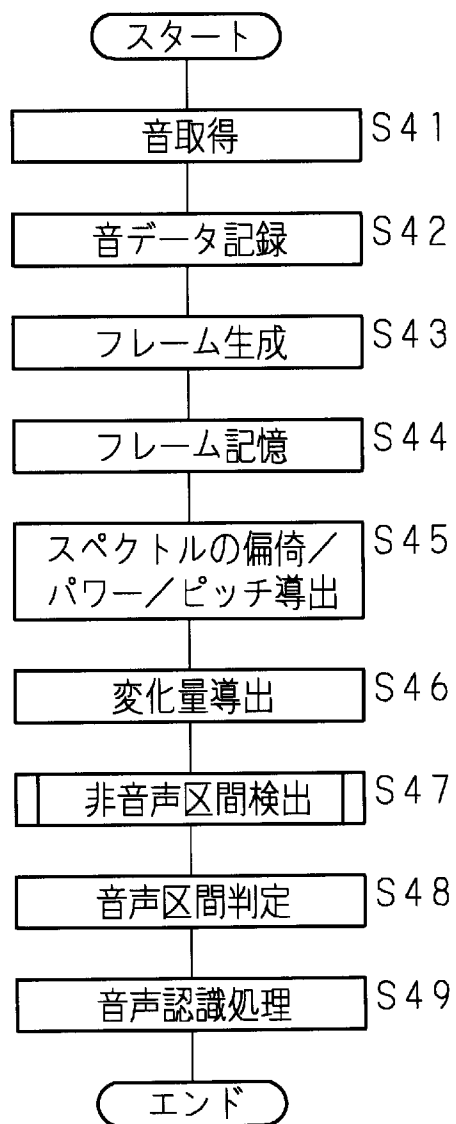
[図9]



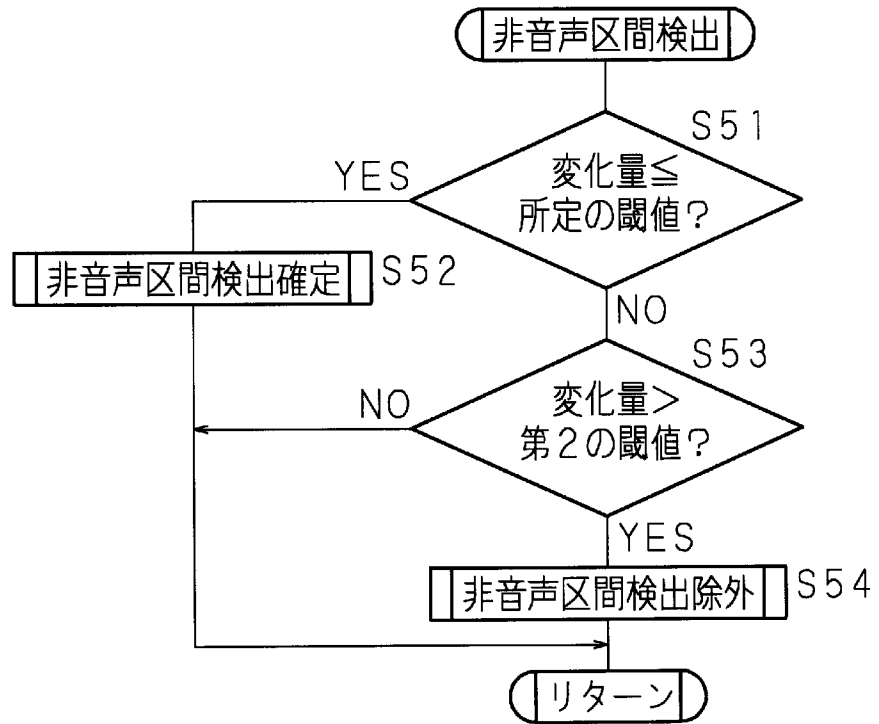
[図10]



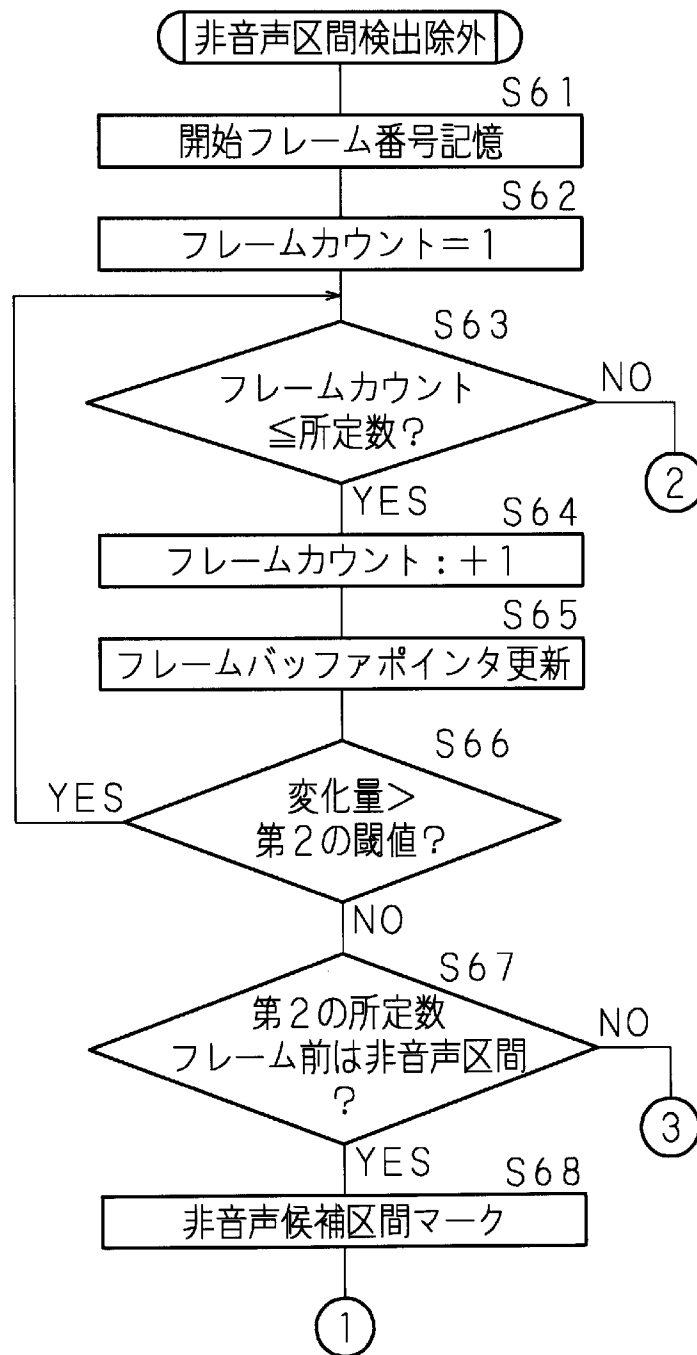
[図11]



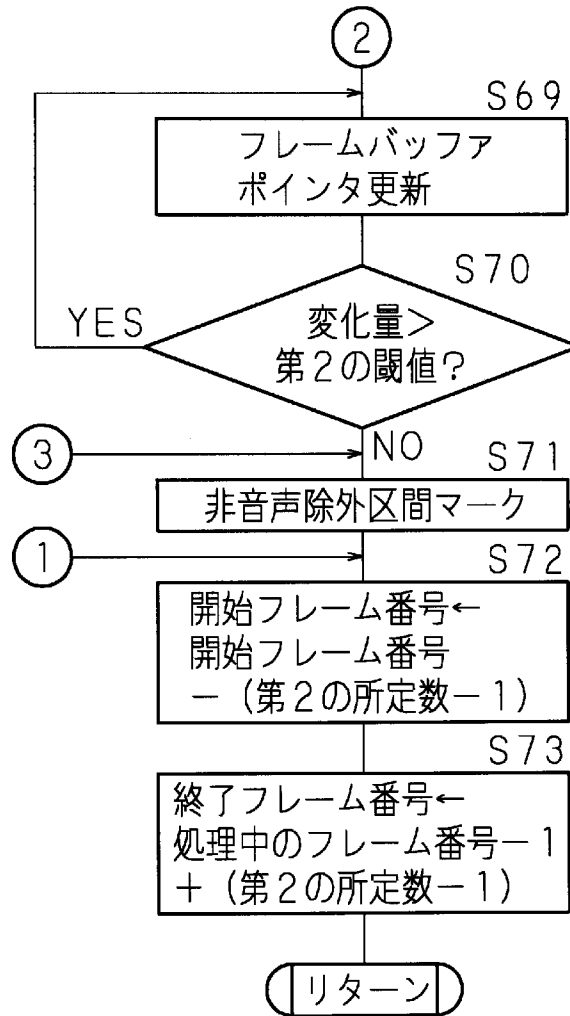
[図12]



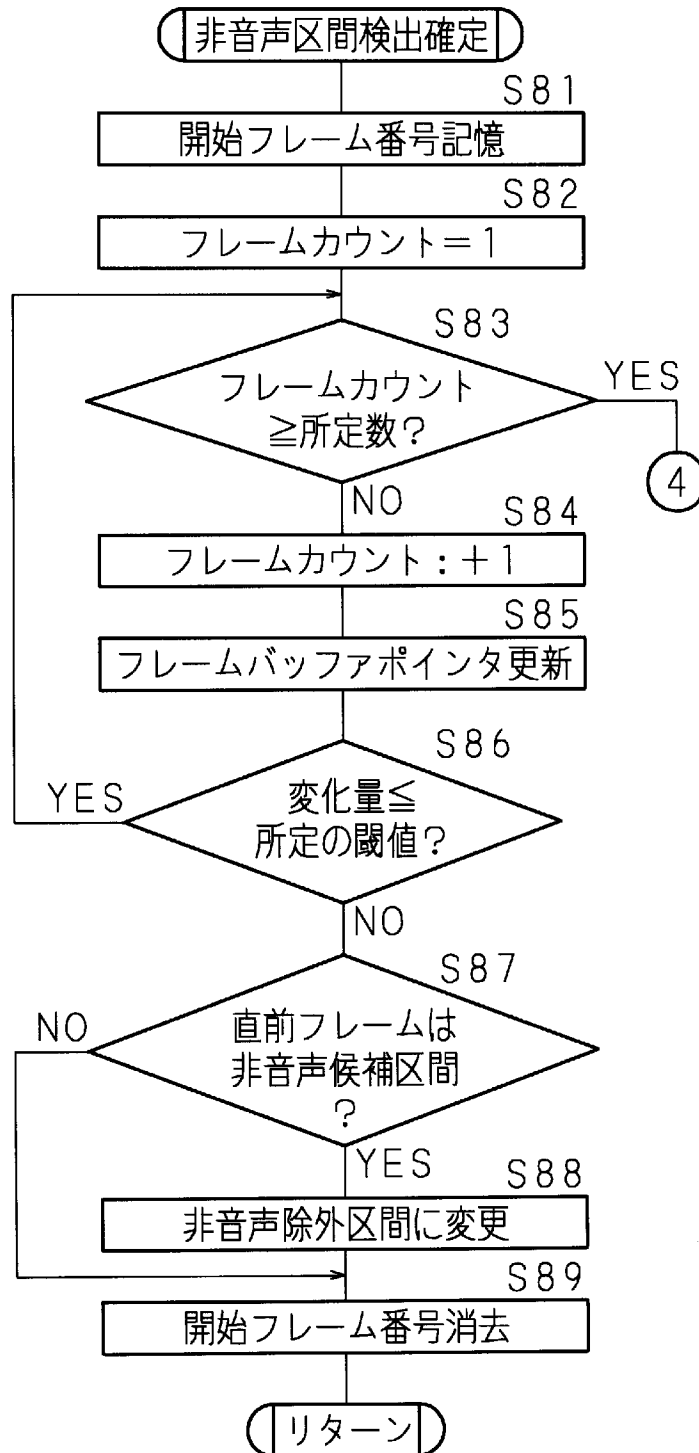
[図13]



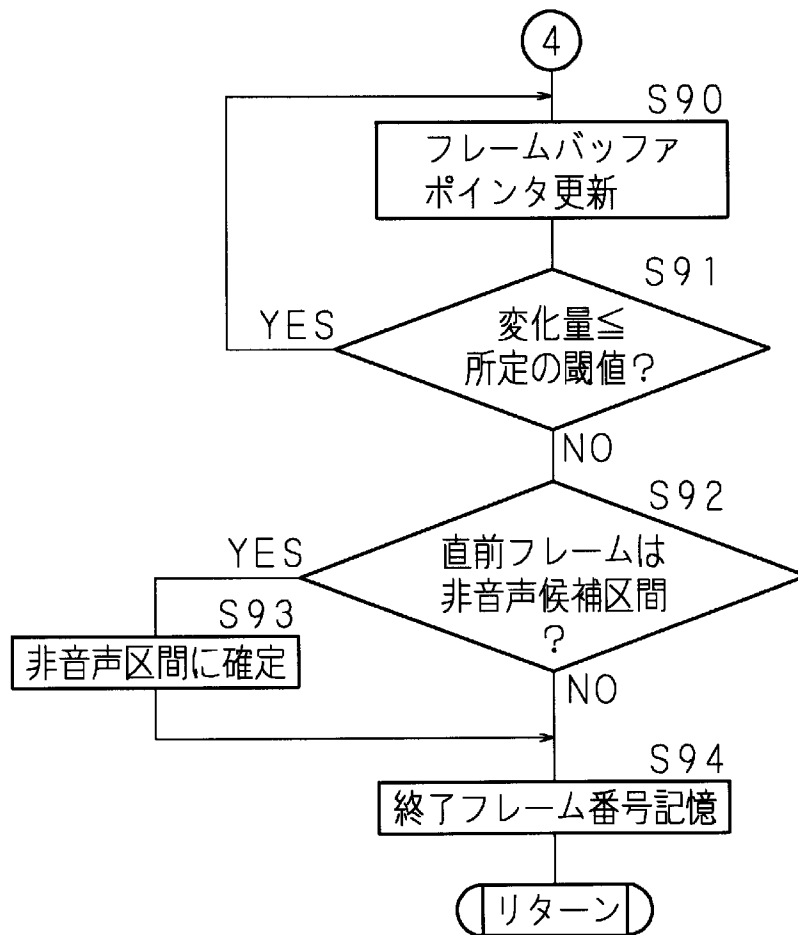
[図14]



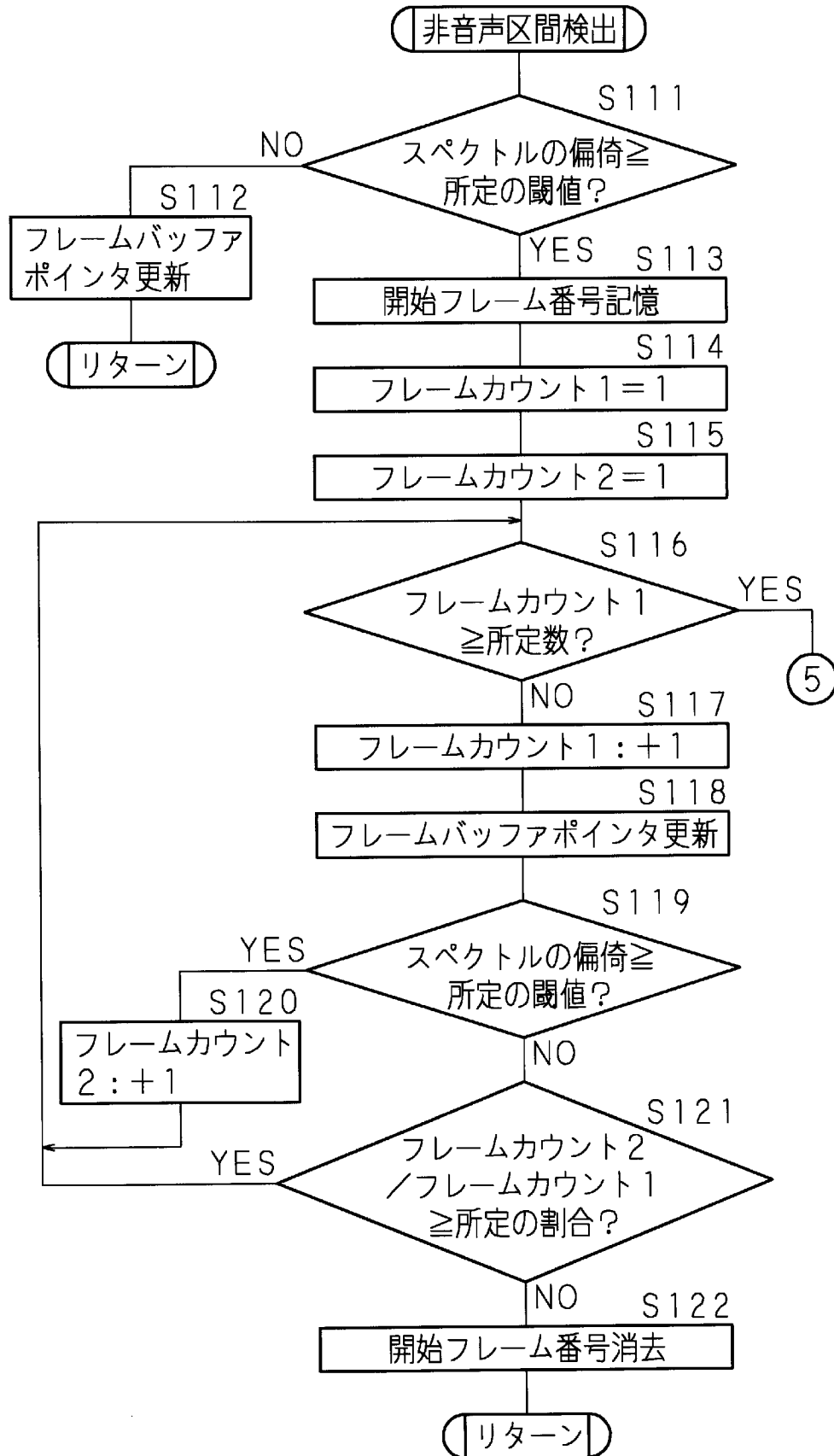
[図15]



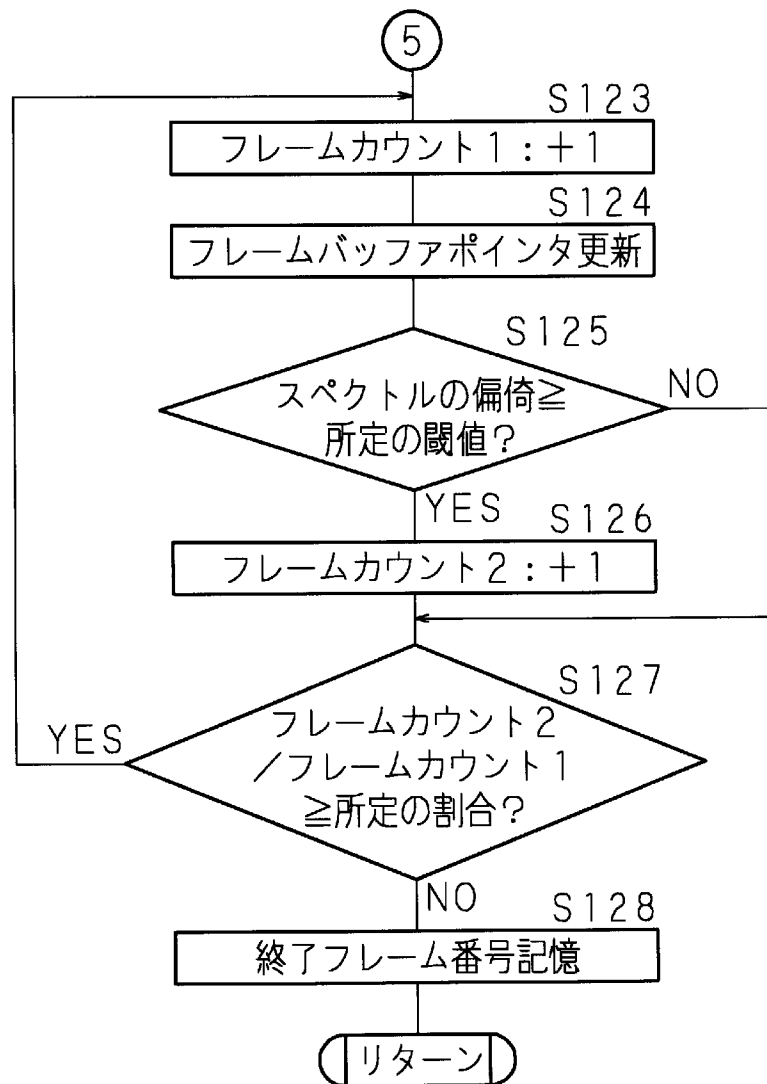
[図16]



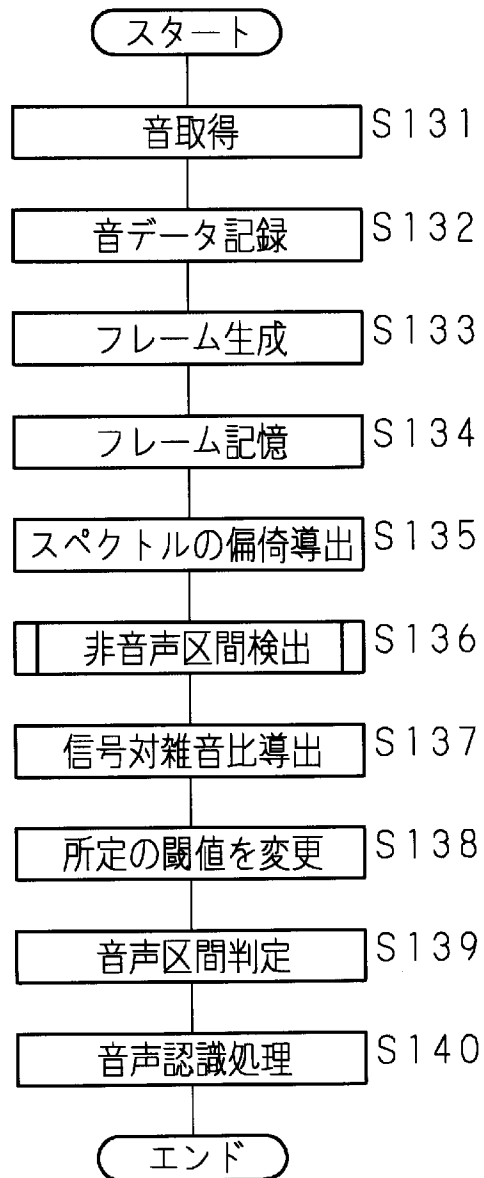
[図17]



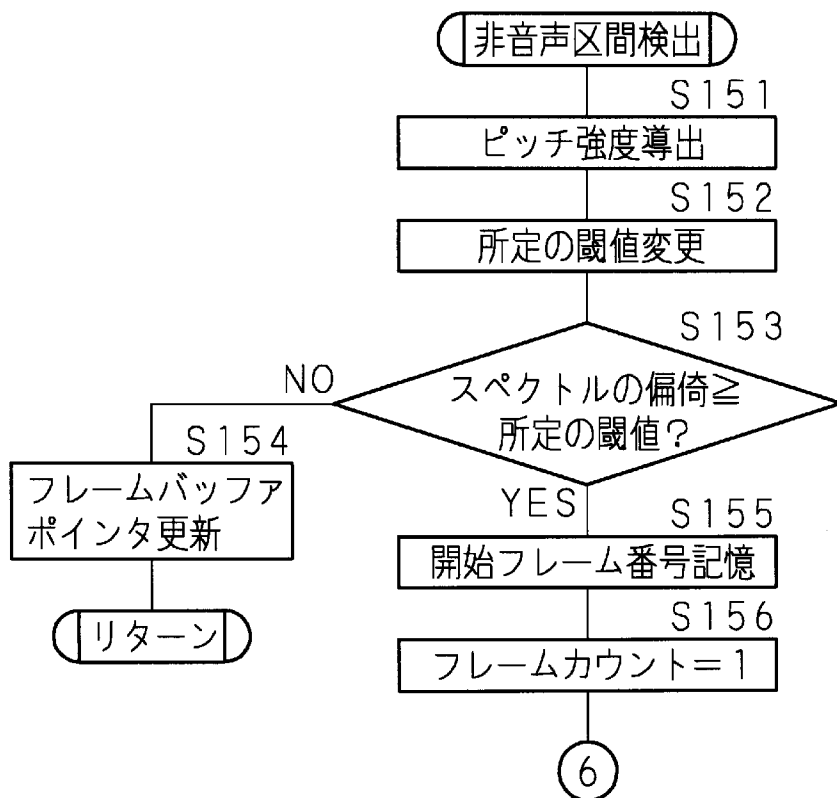
[図18]



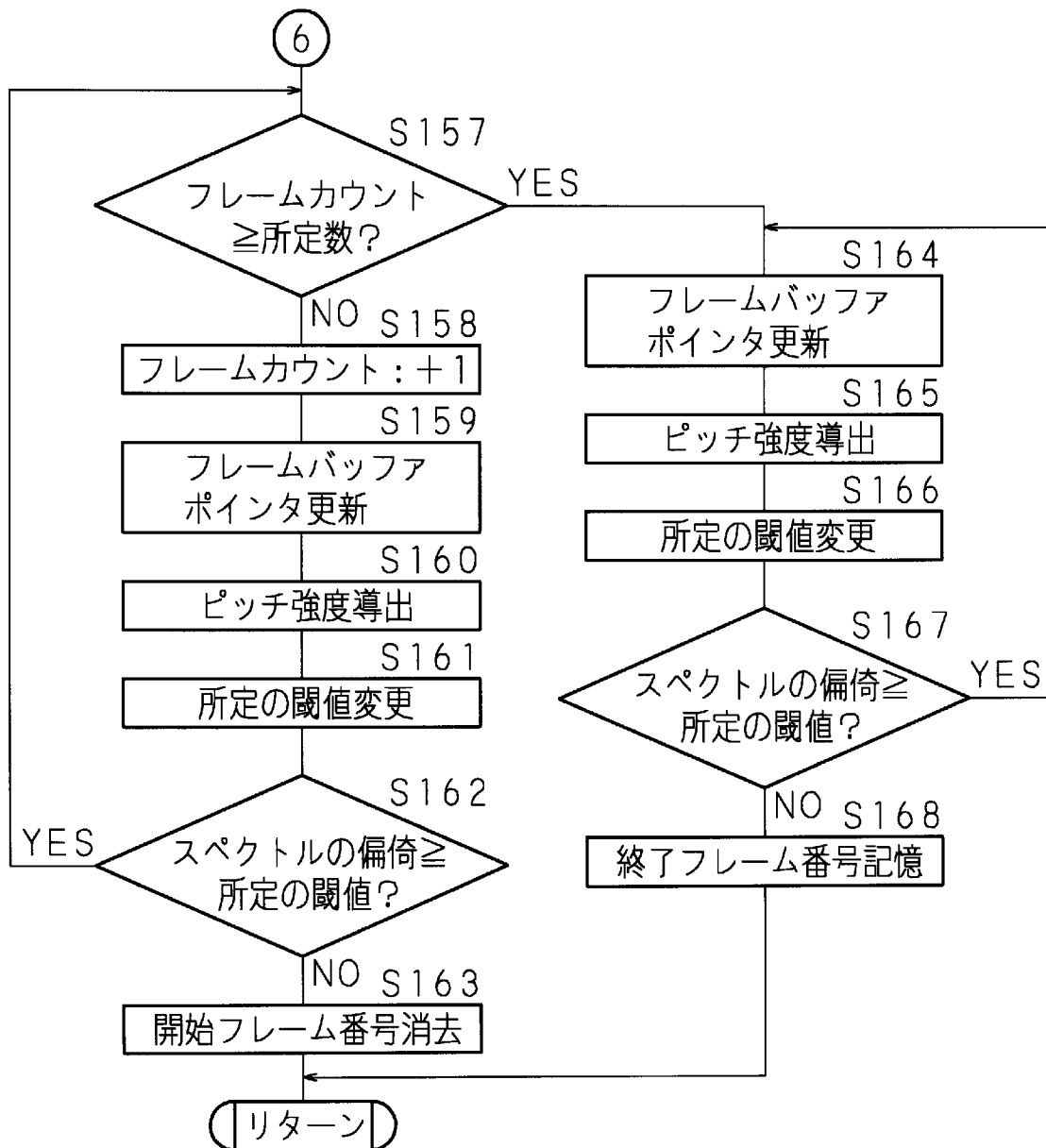
[図19]



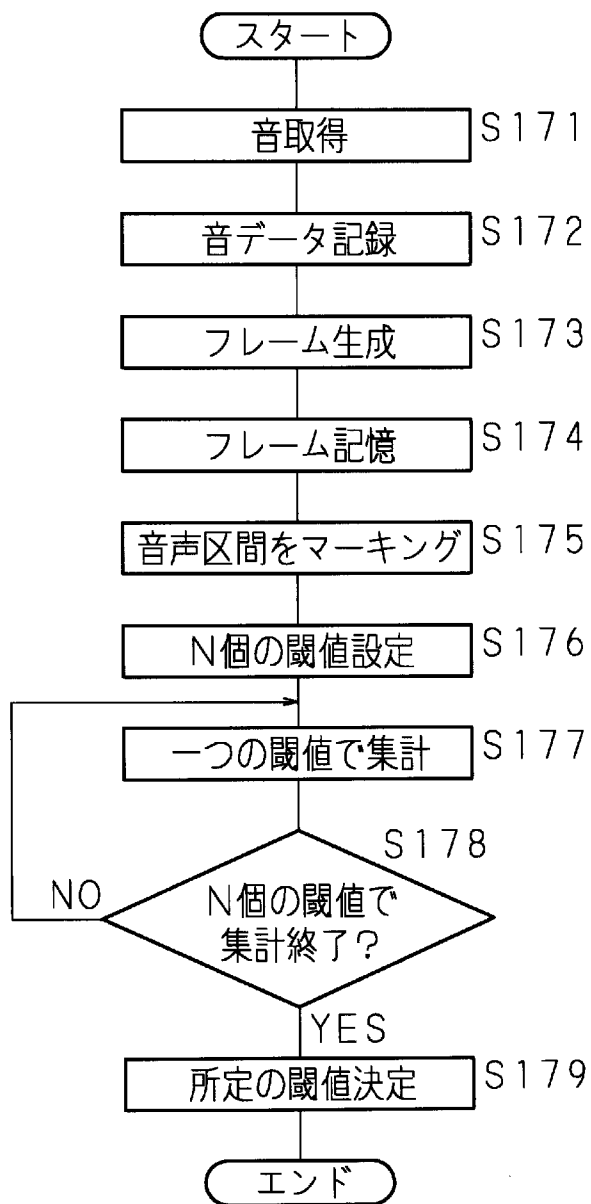
[図20]



[図21]



[図22]



INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2007/074274

A. CLASSIFICATION OF SUBJECT MATTER
G10L11/00(2006.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G10L11/00-11/02, 15/04

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho	1922-1996	Jitsuyo Shinan Toroku Koho	1996-2008
Kokai Jitsuyo Shinan Koho	1971-2008	Toroku Jitsuyo Shinan Koho	1994-2008

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X Y	JP 9-152894 A (Denso Corp.), 10 June, 1997 (10.06.97), Full text; all drawings & US 5937375 A	1, 3, 8, 11, 13 2, 4, 9
Y	JP 2001-350488 A (NEC Corp.), 21 December, 2001 (21.12.01), Full text; all drawings & US 2002/0007270 A1 & US 2006/0271363 A1 & EP 1160763 A2	2, 4, 9
A	JP 6-83391 A (Matsushita Electric Industrial Co., Ltd.), 25 March, 1994 (25.03.94), Full text; all drawings (Family: none)	1-15

Further documents are listed in the continuation of Box C. See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier application or patent but published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search 25 March, 2008 (25.03.08)	Date of mailing of the international search report 08 April, 2008 (08.04.08)
--	---

Name and mailing address of the ISA/ Japanese Patent Office	Authorized officer
Facsimile No.	Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2007/074274

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2005-156887 A (Matsushita Electric Works, Ltd.), 16 June, 2005 (16.06.05), Full text; all drawings (Family: none)	1-15
A	JP 2006-209069 A (Advanced Telecommunication Research Institute International), 10 June, 2006 (10.06.06), Full text; all drawings (Family: none)	1-15
A	JP 2007-233267 A (National Institute of Advanced Industrial Science and Technology), 13 September, 2007 (13.09.07), Full text; all drawings (Family: none)	1-15

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2007/074274

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

The matter common to the inventions of claims 1-15 is that "frames having predetermined time lengths are generated from sound data obtained by sampling sound and a non-speech section including a frame not containing speech data based on a speech uttered by a person is detected". However, the matter cannot be a special technical feature within the meaning of PCT Rule 13.2, second sentence since it is disclosed in, for example, JP 9-152894 A (Denso Corp.).

No other common matter considered as a special technical feature can be seen in claims 1-15.

1. As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest
the

- The additional search fees were accompanied by the applicant's protest and, where applicable, payment of a protest fee.
- The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

A. 発明の属する分野の分類 (国際特許分類 (IPC))
 Int.Cl. G10L11/00(2006.01)i

B. 調査を行った分野
 調査を行った最小限資料 (国際特許分類 (IPC))
 Int.Cl. G10L11/00-11/02, 15/04

最小限資料以外の資料で調査を行った分野に含まれるもの
 日本国実用新案公報 1922-1996年
 日本国公開実用新案公報 1971-2008年
 日本国実用新案登録公報 1996-2008年
 日本国登録実用新案公報 1994-2008年

国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)
 JSTPlus(JDreamII)

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
X Y	JP 9-152894 A (株式会社デンソー) 1997. 06. 10, 全文, 全図 & US 5937375 A	1, 3, 8, 11, 13 2, 4, 9
Y	JP 2001-350488 A (日本電気株式会社) 2001. 12. 21, 全文, 全図 & US 2002/0007270 A1 & US 2006/0271363 A1 & EP 1160763 A2	2, 4, 9

C欄の続きにも文献が列挙されている。 パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー
 「A」特に関連のある文献ではなく、一般的な技術水準を示すもの
 「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
 「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)
 「O」口頭による開示、使用、展示等に言及する文献
 「P」国際出願日前で、かつ優先権の主張の基礎となる出願日の後に公表された文献
 「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの
 「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
 「Y」特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
 「&」同一パテントファミリー文献

国際調査を完了した日 25. 03. 2008	国際調査報告の発送日 08. 04. 2008
----------------------------	----------------------------

国際調査機関の名称及びあて先 日本国特許庁 (ISA/J P) 郵便番号100-8915 東京都千代田区霞が関三丁目4番3号	特許庁審査官 (権限のある職員) 菊池 智紀	5 Z	3 3 5 2
	電話番号 03-3581-1101 内線 3541		

C (続き) . 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
A	JP 6-83391 A (松下電器産業株式会社) 1994. 03. 25, 全文, 全図 (ファミリーなし)	1-15
A	JP 2005-156887 A (松下電工株式会社) 2005. 06. 16, 全文, 全図 (ファミリーなし)	1-15
A	JP 2006-209069 A (株式会社国際電気通信基礎技術研究所) 2006. 06. 10, 全文, 全図 (ファミリーなし)	1-15
A	JP 2007-233267 A (独立行政法人産業技術総合研究所) 2007. 09. 13, 全文, 全図 (ファミリーなし)	1-15

第II欄 請求の範囲の一部の調査ができないときの意見（第1ページの2の続き）

法第8条第3項（PCT17条(2)(a)）の規定により、この国際調査報告は次の理由により請求の範囲の一部について作成しなかった。

1. 請求の範囲 _____ は、この国際調査機関が調査をすることを要しない対象に係るものである。つまり、
2. 請求の範囲 _____ は、有意義な国際調査をすることができる程度まで所定の要件を満たしていない国際出願の部分に係るものである。つまり、
3. 請求の範囲 _____ は、従属請求の範囲であってPCT規則6.4(a)の第2文及び第3文の規定に従って記載されていない。

第III欄 発明の単一性が欠如しているときの意見（第1ページの3の続き）

次に述べるようにこの国際出願に二以上の発明があるところの国際調査機関は認めた。

請求の範囲1-15に共通する事項は、「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」ことであるが、当該事項は、例えば JP 9-152894 A（株式会社デンソー）にも記載されているから、PCT規則13.2の第2文の意味において、特別な技術的特徴とは認められない。

そして、請求の範囲1-15に、特別な技術的特徴と考えられる他の共通な事項を見出すことはできない。

1. 出願人が必要な追加調査手数料をすべて期間内に納付したので、この国際調査報告は、すべての調査可能な請求の範囲について作成した。
2. 追加調査手数料を要求するまでもなく、すべての調査可能な請求の範囲について調査することができたので、追加調査手数料の納付を求めなかった。
3. 出願人が必要な追加調査手数料を一部のみしか期間内に納付しなかったので、この国際調査報告は、手数料の納付のあった次の請求の範囲のみについて作成した。
4. 出願人が必要な追加調査手数料を期間内に納付しなかったので、この国際調査報告は、請求の範囲の最初に記載されている発明に係る次の請求の範囲について作成した。

追加調査手数料の異議の申立てに関する注意

- 追加調査手数料及び、該当する場合には、異議申立手数料の納付と共に、出願人から異議申立てがあった。
- 追加調査手数料の納付と共に出願人から異議申立てがあったが、異議申立手数料が納付命令書に示した期間内に支払われなかった。
- 追加調査手数料の納付はあったが、異議申立てはなかった。