(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2007 年2 月15 日 (15.02.2007)

PCT

(10) 国際公開番号
WO 2007/017993 A1

(51) 国際特許分類:
G10L 11/02 (2006.01) *G10L 15/04* (2006.01)

(21) 国際出願番号: PCT/JP2006/312917

(22) 国際出願日: 2006 年6 月28 日 (28.06.2006)

(25) 国際出願の言語: 日本語

(26) 国際公開の言語: 日本語

(30) 優先権データ:
特願2005-207798 2005 年7 月15 日 (15.07.2005) JP

(71) 出願人 (米国を除く全ての指定国について): ヤマハ株式会社 (YAMAHA CORPORATION) [JP/JP]; 〒4308650 静岡県浜松市中沢町１０番１号 Shizuoka (JP).

(72) 発明者; および

(75) 発明者/出願人 (米国についてのみ): 吉岡 靖雄 (YOSHIOKA, Yasuo).

(74) 代理人: 矢澤 清純 (YAZAWA, Kiyozumi); 〒1050003 東京都港区西新橋一丁目７番１３号 栄光特許事務所 Tokyo (JP).

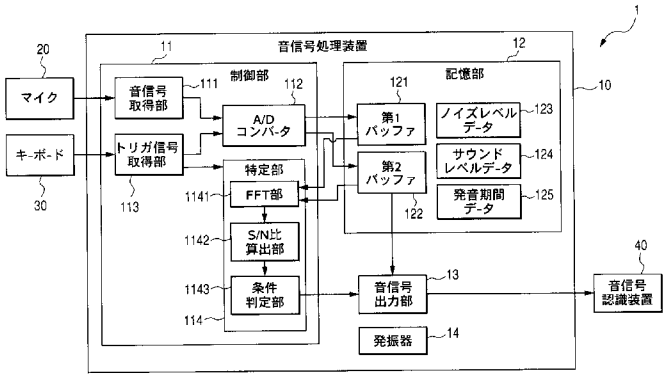
(81) 指定国 (表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(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, 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).

[続葉有]

(54) **Title:** SOUND SIGNAL PROCESSING DEVICE CAPABLE OF IDENTIFYING SOUND GENERATING PERIOD AND SOUND SIGNAL PROCESSING METHOD

(54) **発明の名称:** 発音期間を特定する音信号処理装置および音信号処理方法



20 MICROPHONE
30 KEYBOARD
10 SOUND SIGNAL PROCESSING DEVICE
11 CONTROL UNIT
111 SOUND SIGNAL ACQUIRING SECTION
113 TRIGGER SIGNAL ACQUIRING SECTION
112 A/D CONVERTER
114 IDENTIFYING SECTION
1141 FFT BLOCK
1142 S/N RATIO CALCULATING BLOCK
1143 CONDITION JUDGING BLOCK
12 STORAGE UNIT
121 FIRST BUFFER
122 SECOND BUFFER
123 NOISE LEVEL DATA
124 SOUND LEVEL DATA
125 SOUND GENERATING PERIOD DATA
13 SOUND SIGNAL OUTPUT UNIT
14 OSCILLATOR
40 SOUND SIGNAL RECOGNIZING DEVICE

(57) **Abstract:** A sound generating period of a sound signal can be identified with high accuracy even in a situation where the variation of the environment noise is nonpredicatable. The sound in a sound space where a sound signal processing system (1) is placed is always collected by means of a microphone (20) and inputted as a sound signal into a sound signal processing device (10). Before a predetermined operation is made by the user, the sound signal inputted from the microphone (20) is sequentially stored in a first buffer (121). After the predetermined operation, it is sequentially stored in a second buffer (122). An identifying section (114) calculates the S/N ratio assuming the level of the sound signal stored in a first buffer (121) as the level of the environment noise and the level of the sound signal sequentially stored in the second buffer (122) as the level of the sound being currently generated. The identifying section (114) sequentially judges whether or not the calculated S/N ratio satisfies a predetermined condition, thereby identifying the sound generating period of the sound signal.

[続葉有]

WO 2007/017993 A1

添付公開書類:
— 国際調査報告書
— 請求の範囲の補正の期限前の公開であり、補正書受領 の際には再公開される.

(57) 要約: 環境雑音の変化が予測不可能な状況においても、高い精度で音信号における発音期間の特定を可能とする。音信号処理システム１が置かれた音空間内の音は、常時、マイク２０により収音され、音信号として音信号処理装置１０に入力されている。ユーザにより所定の操作が行われる前は、マイク２０から入力された音信号は第１バッファ１２１に順次格納され、当該所定の操作が行われた後は、第２バッファ１２２に順次格納される。特定部１１４は、第１バッファ１２１に格納されている音信号のレベルを環境雑音のレベルとし、第２バッファ１２２に順次格納される音信号のレベルを現時点で発音されている音のレベルとして、Ｓ／Ｎ比を算出する。特定部１１４は算出したＳ／Ｎ比が所定の条件を満たすか否かを順次判定することにより、音信号における発音期間を特定する。

明 細 書

発音期間を特定する音信号処理装置および音信号処理方法

技術分野

[0001] 本発明は、音信号から発音期間の音を示す部分を特定する技術に関する。

背景技術

[0002] 音声認識やピッチ検出等の処理においては、発音期間、すなわち音声や楽器音が発音されている期間と、非発音期間、すなわち音声や楽器音が発音されていない期間とを区別することが必要である。なぜなら、非発音期間においても通常の音空間には必ず環境雑音が存在するため、仮に発音期間と非発音期間の区別を行うことなく全ての期間において音声認識やピッチ検出等の処理を行うと、非発音期間において環境雑音に基づき誤った処理の結果が得られる可能性があるためである。また、本来処理が不要である非発音期間の音に関し音声認識やピッチ検出等の処理を行うことは無意味であり、処理装置のリソースを無駄に消費する等の観点から好ましくない。

[0003] 音信号における発音期間と非発音期間を区別する方法としては、取得された音信号のS／N(Signal－Noise)比が予め定められたS／N比の閾値を上回る期間を発音期間として特定する方法が広く用いられている。しかしながら、非発音期間における環境雑音のレベルは音信号の取得される環境において様々に変化する。従って、固定的なノイズレベルを用いたS／N比により発音期間の特定を行うと、環境雑音のレベルが高い環境において取得された音信号においては非発音期間が誤って発音期間と特定されたり、環境雑音のレベルが低い環境において取得された音信号においては発音期間が誤って非発音期間と特定されたりする。

[0004] 上記の問題を解決するために、例えば特許文献1には、音声付映像情報から音声情報を抽出するにあたり、音声付映像情報が示すコンテンツのジャンルに応じて異なるノイズレベルを用いる技術が開示されている。

特許文献1:特開2003－101939号公報

[0005] また、例えば特許文献2には、音信号を所定時間長のフレームに分割し、過去に非発音期間と特定されたフレームの属性値に基づき後続のフレームにおけるS／N比

の算出に用いるノイズレベルを更新する技術が開示されている。

特許文献2:特開2001－265367号公報

発明の開示

発明が解決しようとする課題

[0006] ところで、ユーザの本人認証を発声により行う端末装置がある。そのような端末装置においては、ユーザが収音手段を備えた端末装置に対し所定の発声を行う。端末装置は、当該ユーザの発声を示す音信号から特徴量を抽出し、予め記憶されている正しいユーザの発声に関する特徴量と新たに抽出した特徴量とを比較することにより、当該ユーザが正しいユーザであるか否かを判定する。

[0007] 上記のような場合、端末装置は収音手段により取得する音信号のうち、ユーザが発声を行った発音期間を特定する必要がある。ただし、本人認証が行われる際の音空間における環境雑音のレベル等は様々に変化するため、固定的なノイズレベルを用いたS／N比により発音期間の特定を行うと必ずしも正しい結果が得られるとは限らない。また、環境雑音のレベルがどのように変化するかを予め予測することは容易ではないため、特許文献1に開示されるように予めノイズレベルを変更するための基準を与えることも困難である。

[0008] また、特許文献2に開示されるような技術を用いる場合、まず過去のフレームに関し何らかの方法で非発音期間であるか否かの判定を行う必要があり、その判定において用いるノイズレベルを如何に与えるかが問題となる。すなわち、ノイズレベルの初期値が不適当であると発音期間の特定結果の精度が低くなる。

[0009] なお、ユーザによる楽器の演奏音のピッチ検出を行うピッチ検出装置等においても、上述した音声による本人認証を行う端末装置と同様の課題がある。

[0010] 上記の状況に鑑み、本発明は、環境雑音の変化が予測不可能な状況においても、高い精度で音信号における発音期間の特定を可能とする音信号処理装置及び音信号処理方法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

[0011] 上記課題を達成するために、本発明は、継続的に音信号を取得する音信号取得手段と、現時点を終点とする所定期間において前記音信号取得手段により取得され

た音信号を記憶する記憶手段と、トリガ信号を取得するトリガ信号取得手段と、前記トリガ信号が取得された時点後に前記音信号取得手段により取得された音信号を用いてサウンドレベルの指標値を算出し、前記トリガ信号取得手段によりトリガ信号が取得された時点において前記記憶手段に記憶されている音信号を用いてノイズレベルの指標値を算出し、前記サウンドレベルの指標値を前記ノイズレベルの指標値で除すことによりS／N比を算出し、前記S／N比が所定の条件を満たすか否かを判定することにより、前記トリガ信号が取得された時点後に前記音信号取得手段により取得された音信号のうち発音期間の音を示す部分を特定する特定手段とを備える音信号処理装置を提供する。

[0012] かかる音信号処理装置によれば、トリガ信号の取得前に取得され記憶されている音信号を環境雑音のみを示す音信号と見なしてS／N比を算出し、当該S／N比に基づき発音期間の特定が行われる結果、高い精度の特定結果が得られる。

[0013] 前記音信号処理装置において、前記トリガ信号取得手段は、ユーザによる所定の操作に応じて操作手段により生成されるトリガ信号を取得するように構成されてもよいし、ユーザに対し発音を促す通知を行う通知手段により前記通知に伴い生成されるトリガ信号を取得するように構成されてもよい。

[0014] また、前記音信号処理装置において、前記特定手段は、前記トリガ信号が取得された時点後に前記音信号取得手段により取得された音信号の所定周波数の成分のパワーを示す指標値および前記トリガ信号取得手段によりトリガ信号が取得された時点において前記記憶手段に記憶されている音信号の所定周波数の成分のパワーを示す指標値を用いて、前記サウンドレベルの指標値および前記ノイズレベルの指標値をそれぞれ算出するように構成されてもよい。

[0015] また、前記音信号処理装置において、前記特定手段は、前記トリガ信号が取得された時点後に前記音信号取得手段により取得された音信号の振幅値および前記トリガ信号取得手段によりトリガ信号が取得された時点において前記記憶手段に記憶されている音信号の振幅値を用いて、前記サウンドレベルの指標値および前記ノイズレベルの指標値をそれぞれ算出するように構成されてもよい。

[0016] また、前記音信号処理装置において、前記特定手段は、前記トリガ信号が取得さ

れた時点後に前記音信号取得手段により取得された音信号を所定時間長ごとに分割して得られる複数のフレームの各々について前記S／N比を算出し、当該S／N比が所定の条件を満たすフレームの開始時点を前記発音期間の開始時点として特定するように構成されてもよい。そのような態様において、前記特定手段は、所定のフレームに関し算出した前記S／N比が前記所定の条件を満たさない場合、前記記憶手段に記憶されている音信号を当該所定のフレームを用いて更新し、当該所定のフレームの後続のフレームについて前記S／N比を算出するときに、当該更新後の前記記憶手段に記憶されている音信号を用いるように構成されてもよい。

[0017] また、前記音信号処理装置において、前記特定手段は、前記トリガ信号が取得された時点後に前記音信号取得手段により取得された音信号を所定時間長ごとに分割して得られる複数のフレームの各々について前記S／N比を算出し、当該S／N比が所定の条件を満たすフレームの終了時点を前記発音期間の終了時点として特定するように構成されてもよい。

[0018] また、前記音信号処理装置において、前記特定手段は、前記記憶手段に記憶されている音信号を所定時間長ごとに分割して得られる複数のフレームの各々について所定の属性値を算出し、算出した属性値が所定の条件を満たすフレームを前記S／N比の算出に用いないように構成されてもよい。

[0019] また、本発明は、上記の音信号処理装置により行われる処理をコンピュータに実行させるプログラムを提供する。

また、本発明は、継続的に音信号を取得し、現時点を終点とする過去の所定期間において取得した音信号を記憶し、トリガ信号を取得し、前記トリガ信号を取得した時点後に取得した音信号を用いてサウンドレベルの指標値を算出し、前記トリガ信号を取得した時点において記憶している音信号を用いてノイズレベルの指標値を算出し、前記サウンドレベルの指標値を前記ノイズレベルの指標値で除すことによりS／N比を算出し、前記S／N比が所定の条件を満たすか否かを判定し、前記判定処理に基づいて、前記トリガ信号を取得した時点後に取得した音信号のうち発音期間の音を示す部分を特定する音信号処理方法を提供する。

また、本発明の音信号処理方法はさらに、ユーザの操作に応じて所定の信号を生

成し、前記トリガ信号取得処理において、前記ユーザによる所定の操作に応じて前記信号生成処理により生成されるトリガ信号を取得する。

また、本発明の音信号処理方法はさらに、ユーザに対し発音を促す通知を行うとともに、前記通知に伴いトリガ信号を生成し、前記トリガ信号取得処理において、前記通知処理により生成されたトリガ信号を取得する。

また、本発明の音信号処理方法において、前記特定処理は、前記トリガ信号が取得された時点後に前記音信号取得処理により取得された音信号の所定周波数の成分のパワーを示す指標値および前記トリガ信号取得手段によりトリガ信号が取得された時点において前記記憶されている音信号の所定周波数の成分のパワーを示す指標値を用いて、前記サウンドレベルの指標値および前記ノイズレベルの指標値をそれぞれ算出する。

また、本発明の音信号処理方法において、前記特定処理は、前記トリガ信号が取得された時点後に前記音信号取得処理により取得された音信号の振幅値および前記トリガ信号取得処理によりトリガ信号が取得された時点において前記記憶されている音信号の振幅値を用いて、前記サウンドレベルの指標値および前記ノイズレベルの指標値をそれぞれ算出する。

また、本発明の音信号処理方法において、前記特定処理は、前記トリガ信号が取得された時点後に前記音信号取得処理により取得された音信号を所定時間長ごとに分割して得られる複数のフレームの各々について前記S／N比を算出し、当該S／N比が所定の条件を満たすフレームの開始時点を前記発音期間の開始時点として特定する。

また、本発明の音信号処理方法において、前記特定処理は、所定のフレームに関し算出した前記S／N比が前記所定の条件を満たさない場合、前記記憶されている音信号を当該所定のフレームを用いて更新し、当該所定のフレームの後続のフレームについて前記S／N比を算出するときに、当該更新後の前記記憶されている音信号を用いる。

また、本発明の音信号処理方法において、前記特定処理は、前記トリガ信号が取得された時点後に前記音信号取得処理により取得された音信号を所定時間長ごと

に分割して得られる複数のフレームの各々について前記S／N比を算出し、当該S／N比が所定の条件を満たすフレームの終了時点を前記発音期間の終了時点として特定する。

また、本発明の音信号処理方法において、前記特定処理は、前記記憶されている音信号を所定時間長ごとに分割して得られる複数のフレームの各々について所定の属性値を算出し、前記算出した属性値が所定の条件を満たすフレームを前記S／N比の算出に用いない。

発明の効果

[0020] 上記音信号処理装置及び音信号処理方法によれば、トリガ信号の取得前に取得され記憶されている音信号を環境雑音のみを示す音信号と見なしてS／N比を算出し、当該S／N比に基づき発音期間の特定が行われる結果、高い精度の特定結果が得られる。

図面の簡単な説明

[0021] [図1]本発明の実施形態にかかる音信号処理システムの構成を示すブロック図である。

[図2]本発明の実施形態にかかる第1バッファの構成を模式的に示した図である。

[図3]本発明の実施形態にかかる第2バッファの構成を模式的に示した図である。

[図4]本発明の実施形態にかかる周波数帯域の区分を示す図である。

[図5]本発明の実施形態にかかる開始時点の特定処理のフローを示す図である。

[図6]本発明の実施形態にかかる終了時点の特定処理のフローを示す図である。

[図7]本発明の実施形態にかかる発音期間の特定の様子を模式的に示した図である。

符号の説明

[0022] 1 音信号処理システム

10 音信号処理装置

11 制御部

12 記憶部

13 音信号出力部

14 発振器

20 マイク

30 キーボード

40 音信号認識装置

111 音信号取得部

112 A／Dコンバータ

113 トリガ信号取得部

114 特定部

121 第1バッファ

122 第2バッファ

123 ノイズレベルデータ

124 サウンドレベルデータ

125 発音期間データ

1141 FFT部

1142 S／N比算出部

1143 条件判定部

発明を実施するための最良の形態

[0023] ［構成］

図1は本発明の実施形態にかかる音信号処理システム1の構成を示すブロック図である。音信号処理システム1は、取得した音信号における発音期間を特定して特定した発音期間の音信号を出力する音信号処理装置10、置かれた音空間における音を収音し音信号に変換して音信号処理装置10に対し出力するマイク20、複数のキーを有しユーザの当該キーに対する操作に応じて所定の信号を音信号処理装置10に対し出力するキーボード30、音信号処理装置10から出力される音信号の特徴量を抽出し予め記憶している特徴量と比較することにより音信号により示される音声の話者を特定する音信号認識装置40を備えている。

[0024] なお、キーボード30はユーザが音信号処理装置10に対し指示を与える装置の一例であり、マウスポインタ等の他の装置が用いられてもよい。また、音信号認識装置4

0は音信号処理装置10により出力される音信号を利用する装置の一例であり、楽音のピッチを特定する装置等の他の装置が用いられてもよい。

[0025] 音信号処理装置10は、マイク20から音信号を取得して各種処理を行うとともに音信号処理装置10の他の構成部を制御する制御部11、制御部11による各種処理を指示するプログラムおよび制御部11により利用される各種データを記憶するとともに制御部11のワークエリアとして用いられる記憶部12、音信号を音信号認識装置40に対し出力する音信号出力部13、所定時間間隔でクロック信号を生成する発振器14を備えている。なお、音信号処理装置10の各構成部は発振器14により生成されるクロック信号により必要に応じて処理の同期や計時を行う。

[0026] 制御部11は、マイク20から音信号を受け取る音信号取得部111、音信号取得部111が受け取った音信号をアナログ信号からデジタル信号に変換し所定時間長ごとのフレームに区分して記憶部12に順次記憶させるA／D(Analog to Digital)コンバータ112、キーボード30から所定の信号をトリガ信号として受け取るトリガ信号取得部113、トリガ信号取得部113によるトリガ信号の取得をトリガとして記憶部12に順次記憶される音信号における発音期間を特定する特定部114を備えている。

[0027] 記憶部12がA／Dコンバータ112から受け取るフレームには、各フレームを識別するために時系列順にフレーム番号が採番される。以下の説明において、フレーム番号は4桁の整数であり、例えばフレーム番号「0001」のフレームをフレーム「0001」のように呼ぶ。なお、以下の説明において、A／Dコンバータ112により生成されるデジタル信号はPCM(Pulse Code Modulation)形式の音波形データであるものとするが、これに限られない。また、以下の説明においてA／Dコンバータ112により記憶部12に記憶される音信号のフレームの長さは10ミリ秒であるものとするが、これに限られない。

[0028] さらに、特定部114は記憶部12に順次記憶される音信号のフレームの各々に対しFFT(Fast Fourier Transform)アルゴリズムに従った処理を行い当該フレームに含まれる周波数成分を算出するFFT部1141、FFT部1141により算出された周波数成分の振幅を用いてフレームのS／N比を算出するS／N比算出部1142、S／N比算出部1142により算出されたS／N比が所定の条件を満たすか否かを順次判

定することにより発音期間の開始時点および終了時点を特定する条件判定部1143を備えている。S／N比算出部1142および条件判定部1143による具体的な処理内容は後述の動作説明において述べる。

[0029] 記憶部12には、音信号のフレームを一時的に格納するための領域として、第1バッファ121および第2バッファ122が設けられている。第1バッファ121は、音信号処理装置10が動作を開始してからトリガ信号取得部113によりトリガ信号が取得されるまでの間、および前回の発音期間の特定処理がユーザの操作等により終了された後から再びトリガ信号取得部113によりトリガ信号が取得されるまでの間、A／Dコンバータ112により順次生成されるフレームを過去の所定時間長分だけ格納するための領域である。以下、第1バッファ121にフレームの格納が行われる期間を「待機期間」と呼ぶ。また、以下の説明において第1バッファ121に格納可能なフレームは10個、すなわち100ミリ秒分であるものとするが、これに限られない。

[0030] 図2は第1バッファ121の構成を模式的に示した図である。第1バッファ121は10個の領域に分割されており、各領域は「－0010」乃至「－0001」の番号により識別される。以下、例えば番号「－0010」により識別される領域を領域「－0010」のように呼ぶ。第1バッファ121において、領域「－0010」に格納されるフレームが最も古く、領域「－0001」に格納されるフレームが最も新しくなるように、取得順にフレームが各領域に格納される。なお、図2においては、領域「－0010」乃至「－0001」にフレーム「0085」乃至「0094」が各々格納されている様子が例示されている。

[0031] 待機期間中、記憶部12は10ミリ秒間隔でA／Dコンバータ112から新たなフレームを受け取り、FIFO（First－In First－Out）により第1バッファ121の内容を継続的に更新する。なお、図2においては領域「－0010」乃至「－0001」が固定的な位置に描かれているが、記憶部12における各領域の物理的な位置は固定される必要はなく、例えば記憶部12の任意の記憶領域に記憶されたフレームをポインタにより参照することにより、第1バッファ121が実現されてもよい。その場合、ポインタを更新することにより第1バッファ121の内容更新が高速に行われる。

[0032] 第2バッファ122は、トリガ信号取得部113によりトリガ信号が取得された後、ユーザの操作等により発音期間の特定処理が終了されるまでの間、A／Dコンバータ112

により順次生成されるフレームを過去の所定時間長分だけ記憶するための領域である。以下、第2バッファ122にフレームの格納が行われる期間を「判定期間」と呼ぶ。なお、以下の説明において第2バッファ122に格納可能なフレームは6000個、すなわち60秒分であるものとするが、これに限られない。

[0033] 図3は第2バッファ122の構成を模式的に示した図である。第2バッファ122は6000個の領域、すなわち領域「0001」乃至「6000」に分割されている。第2バッファ122において、領域「0001」に格納されるフレームが最も古く、領域「6000」に格納されるフレームが最も新しくなるように、取得順にフレームが各領域に格納される。なお、図3においては、領域「0001」、「0002」、「0003」・・・にフレーム「0095」、「0096」、「0097」・・・が各々格納されている様子が例示されている。また、図3に示される領域「5996」乃至「6000」が空欄となっているのは、図3が判定期間の開始後まだ60秒が経過しておらず、第2バッファ122の末尾付近の領域に未だフレームが格納されていない状態を例示しているためである。

[0034] 判定期間中、記憶部12は10ミリ秒間隔でA／Dコンバータ112から新たなフレームを受け取り、FIFOにより第2バッファ122の内容を継続的に更新する。なお、第2バッファ122に含まれる各領域の物理的な位置が固定される必要はない点は、第1バッファ121の場合と同様である。

[0035] 記憶部12には、さらに、判定期間中にS／N比算出部1142により生成されるノイズレベルデータ123およびサウンドレベルデータ124が一時的に格納される。ノイズレベルデータ123は、トリガ信号取得部113によりトリガ信号が取得された時点において第1バッファ121に記憶されているフレームの振幅に関する属性値を示すデータである。一方、サウンドレベルデータ124は判定期間中に第2バッファ122に順次格納されるフレームの振幅に関する属性値を示すデータである。ノイズレベルデータ123およびサウンドレベルデータ124の具体的内容は後述の動作説明において述べる。

[0036] また、記憶部12には、判定期間中に条件判定部1143により生成される発音期間データ125が一時的に格納される。発音期間データ125は発音期間の先頭のフレーム番号および末尾のフレーム番号を示すデータである。発音期間データ125により、先頭のフレームの開始時点が発音期間の開始時点として特定され、同様に末尾のフ

レームの終了時点が発音期間の終了時点として特定される。なお、発音期間データ125の形式はフレーム番号を用いるものに限られず、例えば発音期間の開始時点および終了時点を時刻データにより特定する等、他に様々なものが考えられる。

[0037] [動作]

続いて、音信号処理システム1の動作を説明する。今、音信号処理システム1のユーザは端末装置(図示略)を利用するために、音信号認識装置40による本人認証を受ける必要があるものとする。

[0038] ユーザは、本人認証を受けるためにキーボード30に対し所定の操作を行い、音信号処理装置10に対し本人認証の処理を指示するが、そのユーザの操作に先立ち、マイク20は常時、音信号処理システム1の配置された音空間の音を示す音信号を音信号処理装置10に対し出力している。音信号処理装置10の音信号取得部111はマイク20から音信号を受け取ると、受け取った音信号を順次、A/Dコンバータ112に引き渡している。そして、A/Dコンバータ112は音信号取得部111から音信号を受け取ると、受け取った音信号をデジタル信号に変換した後、記憶部12に順次引き渡し、フレーム単位で記憶させている。この場合、トリガ信号取得部113はまだトリガ信号を受け取っていないので、待機期間中である。従って、A/Dコンバータ112は記憶部12に対し、送信する音信号を第1バッファ121に格納するように指示している。その結果、第1バッファ121には常に待機期間中における直近の最大10フレーム分の音信号が格納されていることになる。このように第1バッファ121に格納されている音信号は、未だユーザによる発音(発声)が行われていない状態における音空間内の音、すなわち環境雑音の音を示す音信号である。

[0039] 上記の状態において、ユーザがキーボード30に対し所定の操作を行い、音信号処理装置10に対し本人認証の処理を指示すると、キーボード30はユーザの操作に応じてトリガ信号を生成し音信号処理装置10に対し出力する。音信号処理装置10のトリガ信号取得部113はキーボード30からトリガ信号を受け取ると、受け取ったトリガ信号をA/Dコンバータ112および特定部114に送信する。

[0040] A/Dコンバータ112は、トリガ信号取得部113からトリガ信号を受け取ると、その後、生成する音信号を記憶部12に記憶させる際、第2バッファ122に記憶するように指

示する。その結果、第2バッファ122には常に判定期間中における直近の最大6000フレーム分の音信号が格納されていることになる。また、判定期間中において、待機期間中に格納された第1バッファ121の内容は保持されている。

[0041] 一方、特定部114はトリガ信号取得部113からトリガ信号を受け取ると、第2バッファ122に順次格納される音信号における発音期間の特定処理を開始する。まず、FFT部1141は、第1バッファ121に記憶されている直近のフレーム、すなわち領域「－0001」に格納されているフレームに関し、FFT処理を行い、各々のフレームの音信号に含まれる各周波数の成分を示す複素数を算出する。以下、説明のため、第1バッファ121の領域「－0001」に格納されているフレームがフレーム「0094」であるものとする。

[0042] 以下の説明において、FFT部1141はFFT処理により、複数の周波数の成分を示す複素数(R1＋I1i)、(R2＋I2i)、(R3＋I3i)、・・・、(RN＋INi)を算出するものとする。ただし、ここで「i」は虚数単位であり、RnおよびIn(nは1～N、NはFFTbinの数)はそれぞれ実数部および虚数部の数値である。

[0043] FFT部1141は上記のように算出した周波数成分を示す複素数群をS／N比算出部1142に送信する。S／N比算出部1142は、FFT部1141から受け取った複素数群を用いて、複数の周波数帯域に含まれる周波数の複素数群から、予め区分された複数の周波数帯域ごとに、音信号の成分に関する振幅の指標を算出する。以下の説明においては、S／N比算出部1142は図4に示す5つの周波数帯域の各々に関し、以下の(式1)乃至(式5)に従って周波数帯域ごとのパワー:Fm(mは周波数帯域番号)を算出する。ここで、bm:所望帯域の最低周波数に対応するFFTbinの番号、em:所望帯域の最高周波数に対応するFFTbinの番号とする。

[数1]

$$F_1 = \sum_{j=b_1}^{e_1} \sqrt{R_j^2 + I_j^2} \quad \cdots \quad (式1)$$

$$F_2 = \sum_{j=b_2}^{e_2} \sqrt{R_j^2 + I_j^2} \quad \cdots \quad (式2)$$

$$F_3 = \sum_{j=b_3}^{e_3} \sqrt{R_j^2 + I_j^2} \quad \cdots \quad (式3)$$

$$F_4 = \sum_{j=b_4}^{e_4} \sqrt{R_j^2 + I_j^2} \quad \cdots \quad (式4)$$

$$F_5 = \sum_{j=b_5}^{e_5} \sqrt{R_j^2 + I_j^2} \quad \cdots \quad (式5)$$

[0044] 以下、FFT部1141およびS／N比算出部1142によりフレーム「0094」に格納されているフレームに関し上記のように算出されたF1、F2、・・・、F5をF0094_1、F0094_2、・・・、F0094_5のように呼ぶ。

[0045] 続いて、FFT部1141およびS／N比算出部1142は、第1バッファ121の領域「－0002」乃至「－0005」に格納されているフレームの各々に関しても、同様に周波数帯域ごとのパワー、すなわちF0093_1乃至F0093_5、F0092_1乃至F0092_5、F0091_1乃至F0091_5、F0090_1乃至F0090_5を算出する。

[0046] 続いて、S／N比算出部1142は以下の(式6)に従って周波数帯域ごとのノイズレベル：NLm(mは周波数帯域番号)を算出する。ただし、(式6)におけるtはフレーム番号を示し、この場合k＝0090である。

[数2]

$$NL_m = \left(\sum_{t=k}^{k+4} F_{t\_m} \right) \Big/ 5 \quad \cdots \quad (式6)$$

[0047] S／N比算出部1142は上記のようにNL1乃至NL5を算出すると、それらの数値群

を示すデータをノイズレベルデータ123として記憶部12に記憶させる。このように記憶部12に記憶されるノイズレベルデータ123は、環境雑音のレベルを所定の周波数帯域ごとに示すデータである。

[0048] 続いて、FFT部1141は第2バッファ122に新たに格納されたフレーム、すなわちフレーム「0095」に関し、上述した第1バッファ121に格納されたフレームに関するものと同様の処理を行い、周波数成分を示す複素数群を算出する。S／N比算出部1142はFFT部1141によりフレーム「0095」に関し算出された複素数群を用いて、上記(式1)乃至(式5)に従って、周波数帯域ごとのパワー、すなわちF0095_1、F0095_2、・・・、F0095_5を算出する。

[0049] S／N比算出部1142は上記のようにF0095_1乃至F0095_5を算出すると、それらの数値群を示すデータをサウンドレベルデータ124として記憶部12に記憶させる。このように記憶部12に記憶されるサウンドレベルデータ124は、現時点の音空間における音のレベルを所定の周波数帯域ごとに示すデータである。

[0050] S／N比算出部1142は、上記のように記憶部12に記憶したノイズレベルデータ123およびサウンドレベルデータ124を用いて、以下の(式7)に従って、S／N比:SNRを算出する。ただし、(式7)におけるSはサウンドレベルデータ124の算出に用いられたフレーム番号を示し、この場合S＝0095である。

[数3]

$$SNR = \left(\sum_{m=1}^{5} \frac{F_{s\_m}}{NL_m} \right) \Big/ 5 \quad \cdots \quad (式7)$$

[0051] FFT部1141およびS／N比算出部1142は、第2バッファ122に新たなフレームが格納されるごとに、上記(式7)に従い、新たに格納されたフレームに関するSNRを算出する。なお、判定期間中に第1バッファ121に格納されている音信号は変更されないので、第2バッファ122に格納されるフレーム「0096」以降に関するSNRの算出においては、既に記憶部12に記憶されているノイズレベルデータ123が利用される。

[0052] 上記のようにS／N比算出部1142により算出されるSNRは、環境雑音のレベルに対する現時点の音空間における音のレベルの比を示す指標である。従って、ユーザにより発声がなされていない間に算出されるSNRは1の近傍を示し、ユーザにより発声がなされている間に算出されるSNRは1よりかなり大きな数値を示すことになる。そこで、条件判定部1143はS／N比算出部1142により順次算出されるSNRに基づき、発音期間の特定処理を以下のように行う。

[0053] 条件判定部1143による発音期間の特定処理は、発音期間の開始時点を特定するための処理と、発音期間の終了時点を特定するための処理に区分される。図5は開始時点の特定処理のフローを、図6は終了時点の特定処理のフローを、それぞれ示している。

[0054] まず、条件判定部1143は発音期間の特定処理に先立ち、予め以下の定数をパラメータとして記憶部12に記憶している。

(a)開始閾値

(b)開始満了回数

(c)開始猶予回数

(d)終了閾値

(e)終了満了回数

[0055] 開始閾値は、SNRがその値を超えた場合に、そのSNRの算出に用いられたフレームが発音期間中のフレームである可能性が高いことを示す閾値である。以下の説明において、開始閾値＝2．0であるものとする。

[0056] 開始満了回数は、その回数を超えてSNRが開始閾値を超えた場合に、最初に開始閾値を超えたSNRに対応するフレームを発音期間の先頭フレームと判定するための回数である。以下の説明において、開始満了回数＝5であるものとする。

[0057] 開始猶予回数は、いったん発音期間の開始時点の特定処理が開始された後、SNRが開始閾値を超えるか否かの判定がその回数を超えても開始時点の特定がなされない場合に、再度、それ以降のSNRに関し発音期間の開始時点の特定処理をやり直すための回数である。以下の説明において、開始猶予回数＝10であるものとする。

[0058] 終了閾値は、SNRがその値を下回った場合に、そのSNRの算出に用いられたフレームが非発音期間のフレームである可能性が高いことを示す閾値である。以下の説明において、終了閾値＝1．2であるものとする。

[0059] 終了満了回数は、その回数を超えてSNRが終了閾値を下回った場合に、最初に終了閾値を下回ったSNRに対応するフレームを発音期間の末尾フレームと判定するための回数である。以下の説明において、終了満了回数＝15であるものとする。

[0060] 条件判定部1143は、判定期間になると、まず以下の変数を初期化する（ステップS100）。

(f)開始時点データ

(g)試行カウンタ

(h)開始閾値超過カウンタ

[0061] 開始時点データは、発音期間の先頭のフレームのフレーム番号が格納される変数であり、そのフレーム番号の開始時点が発音期間の開始時点を示す。初期化により、開始時点データには例えば未特定値を示す「Null」が代入される。

[0062] 試行カウンタは、ステップS100の初期化処理の後、SNRを開始閾値「2．0」と比較した回数をカウントするカウンタである。また、開始閾値超過カウンタはSNRが開始閾値「2．0」を超えた回数をカウントするカウンタである。初期化により、試行カウンタおよび開始閾値超過カウンタにはそれぞれ「0」が代入される。

[0063] 条件判定部1143はS／N比算出部1142から新たなSNRを取得すると（ステップS101）、試行カウンタに1を加算した後（ステップS102）、ステップS101において取得したSNRが開始閾値「2．0」を超えているか否かを判定する（ステップS103）。SNRが開始閾値「2．0」を超えていない場合（ステップS103：No）、条件判定部1143は試行カウンタが開始猶予回数「10」を超えているか否かを判定する（ステップS104）。

[0064] 試行カウンタが開始猶予回数「10」を超えていない場合（ステップS104：No）、条件判定部1143は処理をステップS101に戻し、次のSNRに関しステップS101以降の処理を繰り返す。

[0065] 一方、開始時点の特定がなされないまま、ステップS101以下の処理が繰り返され、

ステップS102において試行カウンタの値が増加される結果、11になると、ステップS104の判定結果がYesとなる。その場合、条件判定部1143は処理をステップS100に戻し、再度、それ以降のSNRに関し発音期間の開始時点の特定処理をやり直す。

[0066] ユーザは発声を行わない間は、SNRが開始閾値「2．0」を超えないため、条件判定部1143は上記のステップS100乃至ステップS104の処理を繰り返す。そのような状態でユーザが発声を開始すると、ステップS103の判定における結果がYesとなる。その場合、続いて条件判定部1143は開始閾値超過カウンタが「0」であるか否かを判定する（ステップS105）。この場合、開始閾値超過カウンタは「0」であるので（ステップS105：Yes）、条件判定部1143は開始時点データに最後に取得したSNRの算出に用いられたフレーム番号を代入する（ステップS106）。このように代入されるフレーム番号は、発音期間の開始時点を示すフレーム番号の候補である。

[0067] 続いて、条件判定部1143は開始閾値超過カウンタに1を加算した後（ステップS107）、開始閾値超過カウンタが開始満了回数「5」を超えているか否かを判定する（ステップS108）。この場合、開始閾値超過カウンタは「1」でありステップS108の判定結果はNoとなるため、条件判定部1143は処理をステップS101に戻し、新たなSNRに関しステップS101以降の処理を繰り返す。

[0068] 通常、ユーザにより発声が開始され、いったんステップS103における判定結果がYesとなると、後続のフレームに関するSNRに関するステップS103における判定結果もしばらくの間、Yesとなる。ユーザの一続きの発声時間は数秒間に渡り、一方、各フレームの長さは10ミリ秒と短いためである。その結果、再びステップS103の判定結果がYesとなり、ステップS105の判定がなされる場合、既に開始閾値超過カウンタは「1」以上となっているため、その判定結果はNoとなる。その場合、条件判定部1143はステップS106の開始時点データの設定を行わず、ステップS107以降の処理を行う。既に仮設定されている発音期間の開始を示すフレーム番号を変更する必要がないためである。

[0069] ステップS103におけるSNRと開始閾値との比較による判定結果が繰り返しYesとなり、ステップS105以降の処理が繰り返され、ステップS107において開始閾値超過カウンタの値が増加される結果、6になると、ステップS108の判定結果がYesとなる。

その場合、条件判定部1143はその時点で開始時点データに格納されているフレーム番号を発音期間の先頭フレームを示すフレーム番号として確定し、処理を発音期間の終了時点の特定処理のフローに移す。

[0070] ところで、いったんステップS103の判定結果がYesとなっても、例えば一単語の発音における音節間において、短い時間ではあるが発声が途切れたり、発声のレベルが低くなったりする場合がある。そのような場合、ステップS103の判定結果が一時的にNoとなるが、後続のSNRに関するステップS103の判定結果がYesとなるため、それらのフレームは一連の発音期間のフレームと判断されることになる。

[0071] 一方、例えばユーザが物を落とした等により大きな雑音が発生し、発声ではない音によりSNRが一時的に高くなるような場合がある。そのような場合、ステップS103の判定結果が一時的にYesとなるが、後続のSNRに関するステップS103の判定結果がNoとなり、試行カウンタが10を超えた時点で開始時点データに仮設定されていたフレーム番号も初期化されるため、誤ってそのような雑音の発生時点が発音期間の開始時点と判断されることはない。

[0072] 上記のように、発音期間の開始時点の特定処理が完了すると、条件判定部1143は続いて図6に示される発音期間の終了時点の特定処理を開始する。条件判定部1143はまず、以下の変数を初期化する(ステップS200)。

(i)終了時点データ

(j)終了閾値未満カウンタ

[0073] 終了時点データは、発音期間の末尾のフレームのフレーム番号が格納される変数であり、そのフレーム番号の終了時点が発音期間の終了時点を示す。初期化により、終了時点データには例えば未特定値を示す「Null」が代入される。

[0074] 終了閾値未満カウンタはSNRが終了閾値「1．2」を下回った回数をカウントするカウンタである。初期化により、終了閾値未満カウンタには「0」が代入される。

[0075] 条件判定部1143はS／N比算出部1142から新たなSNRを取得すると(ステップS201)、ステップS201において取得したSNRが終了閾値「1．2」を下回っているか否かを判定する(ステップS202)。SNRが終了閾値「1．2」を下回っていない場合(ステップS202:No)、条件判定部1143は続いて当該SNRが開始閾値「2．0」を超えて

いるか否かを判定する(ステップS203)。

[0076] ユーザが発声中においては、SNRが開始閾値「2．0」を超える可能性が高く、その場合、ステップS203の判定結果はYesとなり、条件判定部1143は処理をステップS200に移し、再び、終了時点の特定処理をやり直す。また、ユーザが発声中であって、発音がやや小さくなったりした場合には、SNRが開始閾値「2．0」以下であるが、終了閾値「1．2」以上となることがある。そのような場合(ステップS203:No)、条件判定部1143は変数の初期化は行わず、処理をステップS201に戻し、新たなSNRに関しステップS201以降の処理を繰り返す。

[0077] ユーザが発声を行っている間は、SNRが終了閾値「1．2」を下回らないため、条件判定部1143は上記のステップS200乃至ステップS203の処理を繰り返す。そのような状態でユーザが発声を終了すると、ステップS202の判定における結果がYesとなる。その場合、続いて条件判定部1143は終了閾値未満カウンタが「0」であるか否かを判定する(ステップS204)。この場合、終了閾値未満カウンタは「0」であるので(ステップS204:Yes)、条件判定部1143は終了時点データに最後に取得したSNRの算出に用いられたフレーム番号から1を減じた番号を代入する(ステップS205)。このように代入されるフレーム番号は、発音期間の終了時点を示すフレーム番号の候補である。

[0078] 続いて、条件判定部1143は終了閾値未満カウンタに1を加算した後(ステップS206)、終了閾値未満カウンタが終了満了回数「15」を超えているか否かを判定する(ステップS207)。この場合、終了閾値未満カウンタは「1」でありステップS207の判定結果はNoとなるため、条件判定部1143は処理をステップS201に戻し、新たなSNRに関しステップS201以降の処理を繰り返す。

[0079] その後、ユーザがすぐさま発声を開始しない限り、再びステップS202の判定結果がYesとなる。その場合、ステップS204の判定においては、既に終了閾値未満カウンタが「1」以上となっているため、その判定結果がNoとなる。その場合、条件判定部1143はステップS205の終了時点データの設定を行わず、ステップS206以降の処理を行う。既に仮設定されている発音期間の終了を示すフレーム番号を変更する必要がないためである。

[0080] ステップS202におけるSNRと終了閾値との比較による判定結果が繰り返しYesとなり、ステップS204以降の処理が繰り返され、ステップS206において終了閾値未満カウンタの値が増加される結果、31になると、ステップS207の判定結果がYesとなる。その場合、条件判定部1143はその時点で終了時点データに格納されているフレーム番号を発音期間の末尾フレームを示すフレーム番号として確定し、開始時点データおよび終了時点データの組み合わせを発音期間データ125として記憶部12に記憶させる(ステップS208)。その後、条件判定部1143は音信号出力部13に対し音信号の出力を指示した後(ステップS209)、次の発音期間を特定するために、再び図5に示した発音期間の特定処理に戻る。図7は、上述した特定部114による発音期間の特定の様子を模式的に示した図である。

[0081] 音信号出力部13は、条件判定部1143から音信号の出力を指示されると、第2バッファ122から、発音期間データ125に含まれる開始時点データにより示されるフレーム番号を先頭とし、終了時点データにより示されるフレーム番号を末尾とする一連のフレーム群を読み出し、音信号認識装置40に対し出力する。特定部114は例えば、ユーザにより発音期間の特定処理の終了指示がなされるか、トリガ信号の取得時点の後、発音期間の特定処理の最大時間として予め設定された時間が経過するまでの間、図5および図6に示した判定期間における一連の処理を繰り返す。

[0082] 音信号認識装置40は音信号出力部13から受け取った音信号に対し特徴量抽出を行い、予め記憶されている特徴量と、新たに抽出したユーザの音声に関する特徴量とを比較することにより、ユーザの本人認証を行う。その場合、音信号認識装置40が音信号出力部13から受け取る音信号は、ユーザの発声期間に応じた部分が正しく切り出された音信号であるため、精度の高い本人認証が行われることになる。

[0083] 上述したように、音信号処理システム1によれば、環境雑音の変化が予測不可能な状況においても、ユーザによる本人認証の処理の指示をトリガとして、そのトリガの発生直前に取得された音信号を環境雑音を示す音信号として用いる結果、トリガの発生後の音信号における発音期間の特定を高い精度で行うことが可能となる。

[0084] [変形例]

ところで、上述した実施形態は以下のように様々に変形してもよい。まず、上記説明

においては、トリガ信号はユーザのキーボード30に対する操作に応じて生成されるものとした。それに代えて、例えば、音信号処理装置10が画像や文字でユーザにメッセージを通知するディスプレイや音でユーザにメッセージを通知するサウンドシステム等の通知手段を備えるようにし、ユーザの操作を待つことなく、例えば予め定められた時刻においてそれらの通知手段を用いてユーザに対し発声を促す通知を行うと同時に、トリガ信号を生成しトリガ信号取得部113に該トリガ信号を送信する渡すようにしてもよい。

[0085] また、上記説明においては、複数の周波数帯域ごとにノイズレベルを示すNLmおよびサウンドレベルを示すFmを算出した後、周波数帯域ごとのFm／NLmの平均値を算出することによりSNRを算出するものとした(式1乃至7参照)。それに代えて、例えば、S／N比算出部1142が全周波数帯域に関するNLおよびFを各1つずつ算出した後、SNR＝F／NLとして算出するようにしてもよい。すなわち、周波数帯域の区分数は1であってもよい。

[0086] また、上記説明においては、音信号に対しFFT部1141がFFT処理を施し、S／N比算出部1142が各周波数の振幅を算出することにより、周波数帯域ごとのパワーを示すFmを算出するものとした。それに代えて、例えば、特定部114がFFT部1141の代わりに周波数帯域ごとのバンドパスフィルタを備えるようにし、各バンドパスフィルタにより濾波された音信号の振幅の平均値を上記(式6)および(式7)におけるFmの代わりに用いることにより、SNRを算出するようにしてもよい。

[0087] さらに、FFT部1141やバンドパスフィルタを用いることなく、第1バッファ121に格納されているフレームに含まれる音信号の振幅の平均値で第2バッファ122に格納されているフレームに含まれる音信号の振幅の平均値を単純に除すことにより、SNRを算出するようにしてもよい。

[0088] また、上記説明においては、S／N比算出部1142は各周波数成分のパワーを示すFmを上記(式1)乃至(式5)に従い算出するものとした。それに代えて、例えば、以下の(式8)もしくは(式9)に従いFmを算出するようにしてもよい。ただし、(式9)における「abs()」は()内の数値の絶対値を示す。

[数4]

$$F_{\mathrm{m}} = \sum_{\mathrm{j}} \left(R_j^2 + I_j^2 \right) \quad \cdots \quad (式８)$$

$$F_{\mathrm{m}} = \sum_{j} \left(abs\left(R_j\right) + abs\left(I_j\right) \right) \quad \cdots \quad (式９)$$

[0089] また、上記説明においては、SNRを算出するにあたり、周波数帯域ごとに算出されたFm／NLmを単純平均するものとした(式7参照)。それに代えて、例えば、ユーザにより発音される音が有する割合が高いと予想される周波数成分を含む周波数帯域に関するFm／NLmに相対的に大きな値が設定されたウェイトを用いて、S／N比算出部1142がFm／NLmの加重平均を行うことにより、SNRを算出するようにしてもよい。

[0090] また、上記説明においては、トリガ信号が取得された後に第1バッファ121の内容が変更されることはなく、いったんノイズレベルを示すNLm(式6参照)が算出されると、その後の発音期間の特定処理においてNLmが更新されることはないものとした。それに代えて、例えば、図5のステップS103におけるSNRが開始閾値を超えるか否かの判定結果がNoとなり、そのSNRの算出に用いられたフレームが非発音期間のものであることが確定した時点で、そのフレームを直近の環境雑音を示すフレームとして第1バッファ121に格納することにより、第1バッファ121の内容を更新するようにしてもよい。その場合、FFT部1141およびS／N比算出部1142は更新された第1バッファ121のフレームを用いてNLmを再計算し、その後は再計算されたNLmを用いてSNRの算出を行う。

[0091] また、上記説明においては、ノイズレベルを示すNLm(式6参照)を算出するにあたり、第1バッファ121に格納されている直近の5フレームを固定的に選択するものとした。それに代えて、例えば、第1バッファ121に格納されているフレームの中から異常値を示すフレームを除外し、適当と思われるフレームを選択してNLmの算出に用いるようにしてもよい。具体例を挙げると、FFT部1141は第1バッファ121に格納されている10フレームの全てに関しFFT処理を施す。そして、S／N比算出部1142はそれら10フレームの全てに関し周波数帯域ごとのパワーを示すFmを算出する。そして、S／N比算出部1142はそのようにして算出したFmの平均値から所定の閾値以上に乖

離するFmを異常値として除外し、除外しなかったFmを用いてNLmを算出するようにすればよい。

[0092] また、上記説明においては、第1バッファ121に格納されているフレームの各々に関し算出した周波数帯域ごとのパワーを示すFmを単純平均することによりノイズレベルを示すNLmを算出するものとした(式6参照)。それに代えて、例えば、新しいフレームほど大きなウェイトを与え、S/N比算出部1142が各フレームに関するFmを加重平均することによりNLmを算出するようにしてもよい。

[0093] また、上記説明においては、開始閾値、開始満了回数、開始猶予回数、終了閾値および終了満了回数は予め音信号処理装置10に記憶されているものとしたが、例えば、ユーザの操作に応じてこれらの定数を変更可能としてもよい。

[0094] また、上記説明においては、マイク20、キーボード30および音信号認識装置40は音信号処理装置10とは異なる筐体に配置されているものとしたが、これらの配置は自由に変更可能である。例えば、音信号処理装置10が音信号認識装置40を構成部として備えるようにしてもよい。

[0095] また、音信号処理装置10は、専用のハードウェアにより実現されてもよいし、音信号の入出力が可能な汎用コンピュータにアプリケーションプログラムに従った処理を実行させることにより実現されてもよい。音信号処理装置10が汎用コンピュータにより実現される場合、制御部11は汎用コンピュータが備えるCPU(Central Processing Unit)およびCPUの制御下で動作するDSP(Digital Signal Processor)が、アプリケーションプログラムに含まれる各モジュールに従った処理を同時並行して行うことにより、汎用コンピュータの機能として実現される。

[0096] 本発明を詳細にまた特定の実施態様を参照して説明してきたが、本発明の精神、範囲または意図の範囲を逸脱することなく様々な変更や修正を加えることができることは当業者にとって明らかである。

本発明は、2005年7月15日出願の日本特許出願(特願2000-207798)に基づくものであり、その内容はここに参照として取り込まれる。

産業上の利用可能性

[0097] 本発明の音信号処理装置及び音信号処理方法によれば、トリガ信号の取得前に

取得され記憶されている音信号を環境雑音のみを示す音信号と見なしてS／N比を算出し、当該S／N比に基づき発音期間の特定が行われる結果、高い精度の特定結果が得られる。

請求の範囲

[1] 音信号処理装置は、

継続的に音信号を取得する音信号取得手段と、

現時点を終点とする所定期間において前記音信号取得手段により取得された音信号を記憶する記憶手段と、

トリガ信号を取得するトリガ信号取得手段と、

前記トリガ信号が取得された時点後に前記音信号取得手段により取得された音信号を用いてサウンドレベルの指標値を算出し、前記トリガ信号取得手段によりトリガ信号が取得された時点において前記記憶手段に記憶されている音信号を用いてノイズレベルの指標値を算出し、前記サウンドレベルの指標値を前記ノイズレベルの指標値で除すことによりS／N比を算出し、前記S／N比が所定の条件を満たすか否かを判定することにより、前記トリガ信号が取得された時点後に前記音信号取得手段により取得された音信号のうち発音期間の音を示す部分を特定する特定手段と、

を備える。

[2] 請求項1に記載の音信号処理装置はさらに、ユーザの操作に応じて所定の信号を生成する操作手段を備え、

前記トリガ信号取得手段は、前記ユーザによる所定の操作に応じて前記操作手段により生成されるトリガ信号を取得する。

[3] 請求項1に記載の音信号処理装置はさらに、ユーザに対し発音を促す通知を行うとともに、前記通知に伴いトリガ信号を生成する通知手段を備え、

前記トリガ信号取得手段は、前記通知手段により生成されたトリガ信号を取得する。

[4] 請求項1に記載の音信号処理装置であって、前記特定手段は、前記トリガ信号が取得された時点後に前記音信号取得手段により取得された音信号の所定周波数の成分のパワーを示す指標値および前記トリガ信号取得手段によりトリガ信号が取得された時点において前記記憶手段に記憶されている音信号の所定周波数の成分のパワーを示す指標値を用いて、前記サウンドレベルの指標値および前記ノイズレベルの指標値をそれぞれ算出する。

[5] 請求項1に記載の音信号処理装置であって、前記特定手段は、前記トリガ信号が

取得された時点後に前記音信号取得手段により取得された音信号の振幅値および前記トリガ信号取得手段によりトリガ信号が取得された時点において前記記憶手段に記憶されている音信号の振幅値を用いて、前記サウンドレベルの指標値および前記ノイズレベルの指標値をそれぞれ算出する。

[6] 請求項1に記載の音信号処理装置であって、前記特定手段は、前記トリガ信号が取得された時点後に前記音信号取得手段により取得された音信号を所定時間長ごとに分割して得られる複数のフレームの各々について前記S／N比を算出し、当該S／N比が所定の条件を満たすフレームの開始時点を前記発音期間の開始時点として特定する。

[7] 請求項6に記載の音信号処理装置であって、前記特定手段は、所定のフレームに関し算出した前記S／N比が前記所定の条件を満たさない場合、前記記憶手段に記憶されている音信号を当該所定のフレームを用いて更新し、当該所定のフレームの後続のフレームについて前記S／N比を算出するときに、当該更新後の前記記憶手段に記憶されている音信号を用いる。

[8] 請求項1に記載の音信号処理装置であって、前記特定手段は、前記トリガ信号が取得された時点後に前記音信号取得手段により取得された音信号を所定時間長ごとに分割して得られる複数のフレームの各々について前記S／N比を算出し、当該S／N比が所定の条件を満たすフレームの終了時点を前記発音期間の終了時点として特定する。

[9] 請求項1に記載の音信号処理装置であって、前記特定手段は、前記記憶手段に記憶されている音信号を所定時間長ごとに分割して得られる複数のフレームの各々について所定の属性値を算出し、前記算出した属性値が所定の条件を満たすフレームを前記S／N比の算出に用いない。

[10] 音信号処理方法は、

継続的に音信号を取得し、

現時点を終点とする過去の所定期間において取得した音信号を記憶し、

トリガ信号を取得し、

前記トリガ信号を取得した時点後に取得した音信号を用いてサウンドレベルの指標

値を算出し、

前記トリガ信号を取得した時点において記憶している音信号を用いてノイズレベルの指標値を算出し、

前記サウンドレベルの指標値を前記ノイズレベルの指標値で除すことによりS／N比を算出し、

前記S／N比が所定の条件を満たすか否かを判定し、

前記判定処理に基づいて、前記トリガ信号を取得した時点後に取得した音信号のうち発音期間の音を示す部分を特定する。

[11] 請求項10に記載の音信号処理方法はさらに、ユーザの操作に応じて所定の信号を生成し、

前記トリガ信号取得処理において、前記ユーザによる所定の操作に応じて前記信号生成処理により生成されるトリガ信号を取得する。

[12] 請求項10に記載の音信号処理方法はさらに、ユーザに対し発音を促す通知を行うとともに、前記通知に伴いトリガ信号を生成し、

前記トリガ信号取得処理において、前記通知処理により生成されたトリガ信号を取得する。

[13] 請求項10に記載の音信号処理方法であって、前記特定処理は、前記トリガ信号が取得された時点後に前記音信号取得処理により取得された音信号の所定周波数の成分のパワーを示す指標値および前記トリガ信号取得手段によりトリガ信号が取得された時点において前記記憶されている音信号の所定周波数の成分のパワーを示す指標値を用いて、前記サウンドレベルの指標値および前記ノイズレベルの指標値をそれぞれ算出する。

[14] 請求項10に記載の音信号処理方法であって、前記特定処理は、前記トリガ信号が取得された時点後に前記音信号取得処理により取得された音信号の振幅値および前記トリガ信号取得処理によりトリガ信号が取得された時点において前記記憶されている音信号の振幅値を用いて、前記サウンドレベルの指標値および前記ノイズレベルの指標値をそれぞれ算出する。

[15] 請求項10に記載の音信号処理方法であって、前記特定処理は、前記トリガ信号が

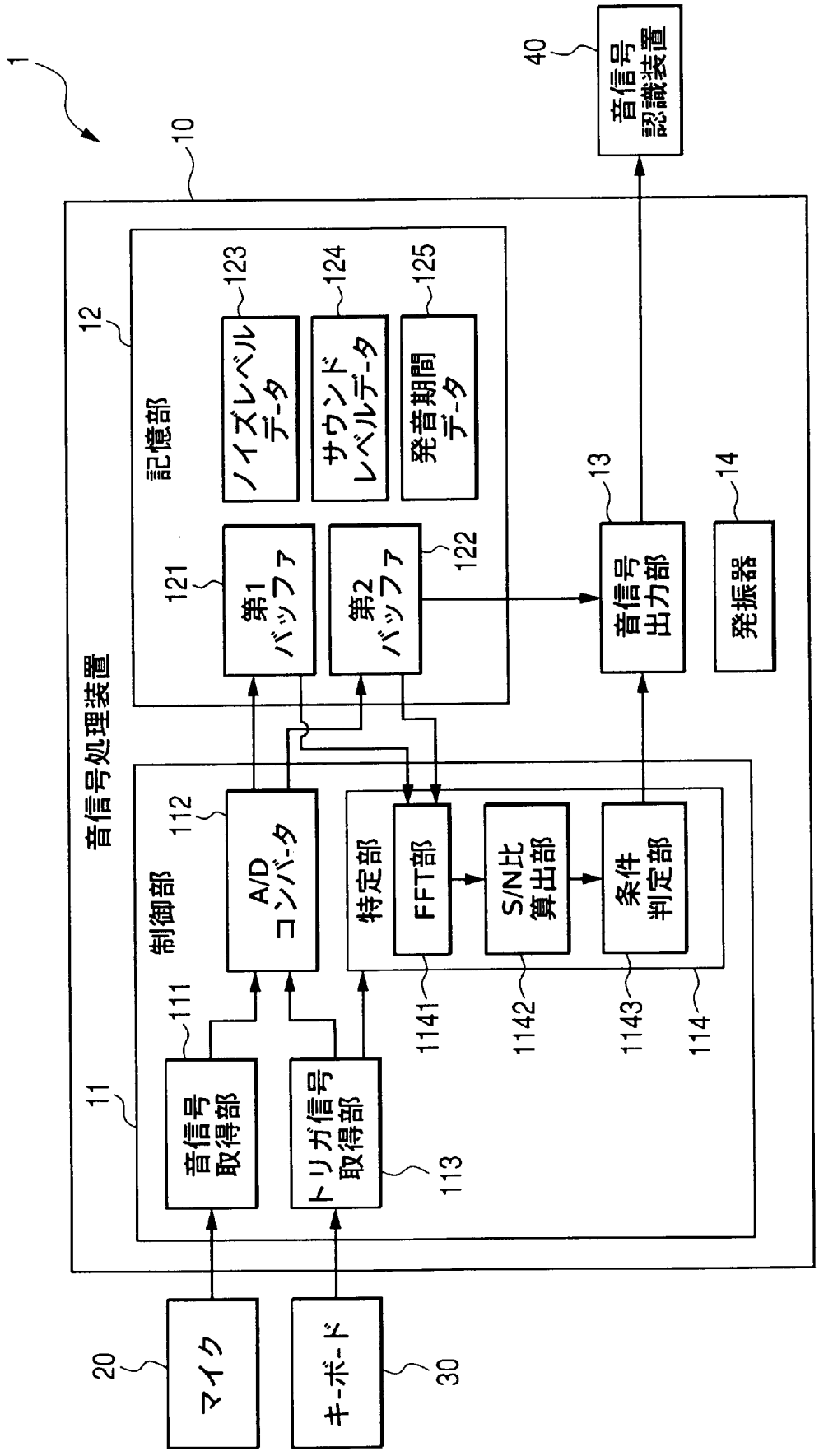
取得された時点後に前記音信号取得処理により取得された音信号を所定時間長ごとに分割して得られる複数のフレームの各々について前記S／N比を算出し、当該S／N比が所定の条件を満たすフレームの開始時点を前記発音期間の開始時点として特定する。

[16] 請求項15に記載の音信号処理方法であって、前記特定処理は、所定のフレームに関し算出した前記S／N比が前記所定の条件を満たさない場合、前記記憶されている音信号を当該所定のフレームを用いて更新し、当該所定のフレームの後続のフレームについて前記S／N比を算出するときに、当該更新後の前記記憶されている音信号を用いる。

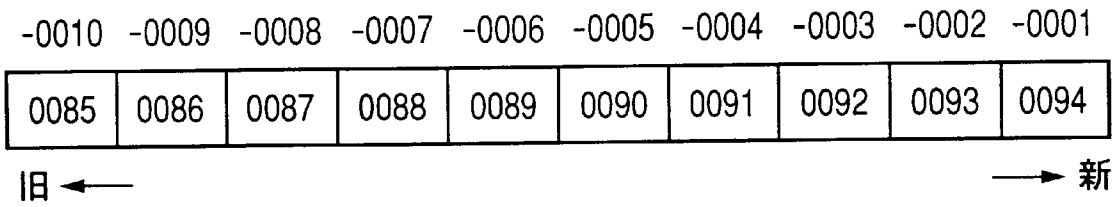
[17] 請求項10に記載の音信号処理方法であって、前記特定処理は、前記トリガ信号が取得された時点後に前記音信号取得処理により取得された音信号を所定時間長ごとに分割して得られる複数のフレームの各々について前記S／N比を算出し、当該S／N比が所定の条件を満たすフレームの終了時点を前記発音期間の終了時点として特定する。

[18] 請求項10に記載の音信号処理方法であって、前記特定処理は、前記記憶されている音信号を所定時間長ごとに分割して得られる複数のフレームの各々について所定の属性値を算出し、前記算出した属性値が所定の条件を満たすフレームを前記S／N比の算出に用いない。

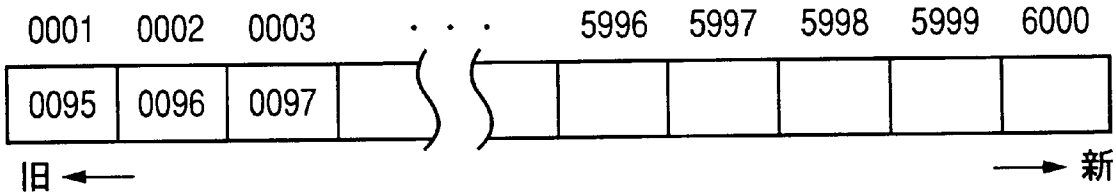
[図1]


1
10
40
音信号認識装置
音信号処理装置
12
記憶部
123
ノイズレベルデータ
124
サウンドレベルデータ
125
発音期間データ
121
第1バッファ
122
第2バッファ
13
音信号出力部
14
発振器
制御部
112
A/Dコンバータ
特定部
FFT部
S/N比算出部
条件判定部
1141
1142
1143
114
11
111
音信号取得部
トリガ信号取得部
113
20
マイク
30
キーボード

[図2]



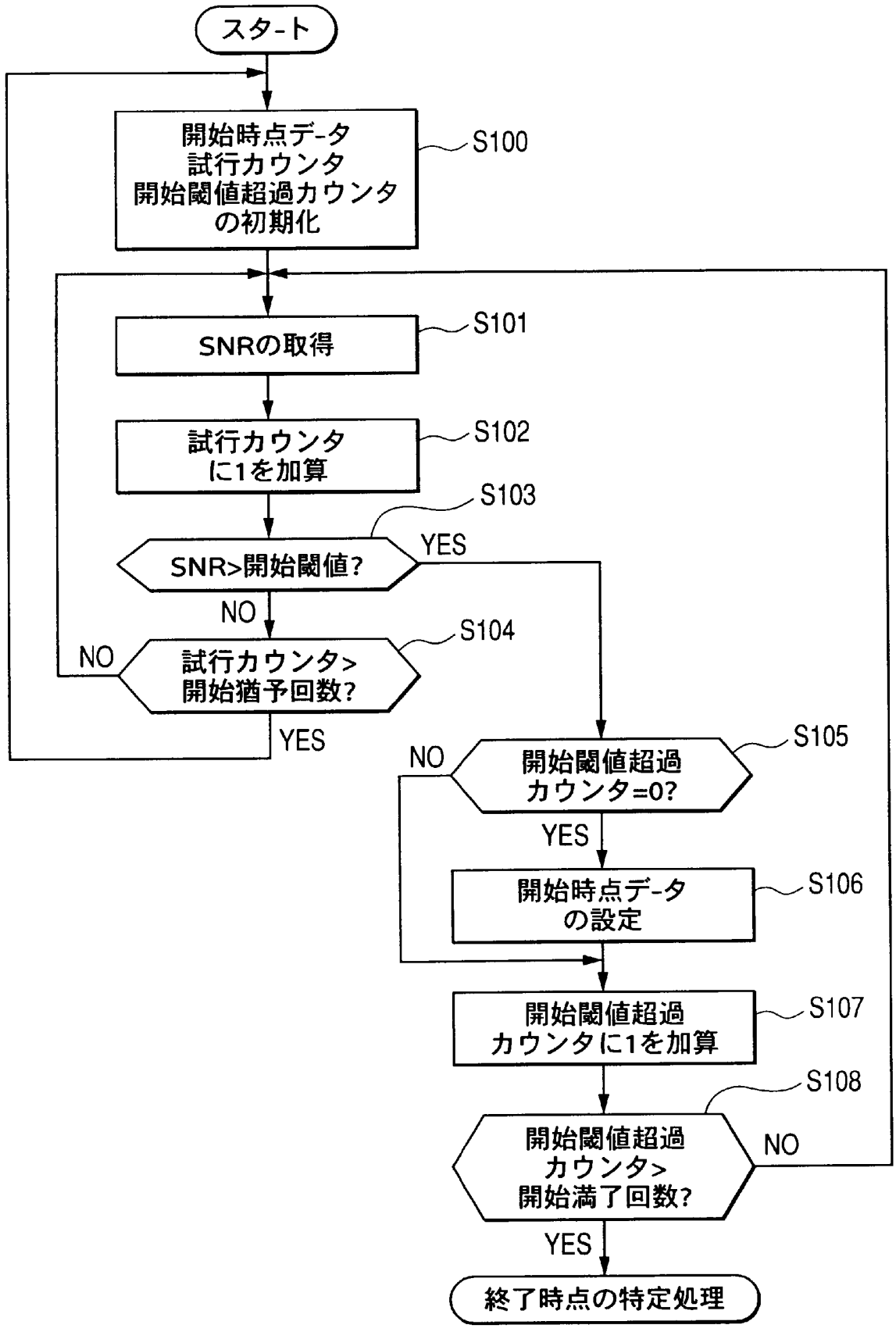
[図3]



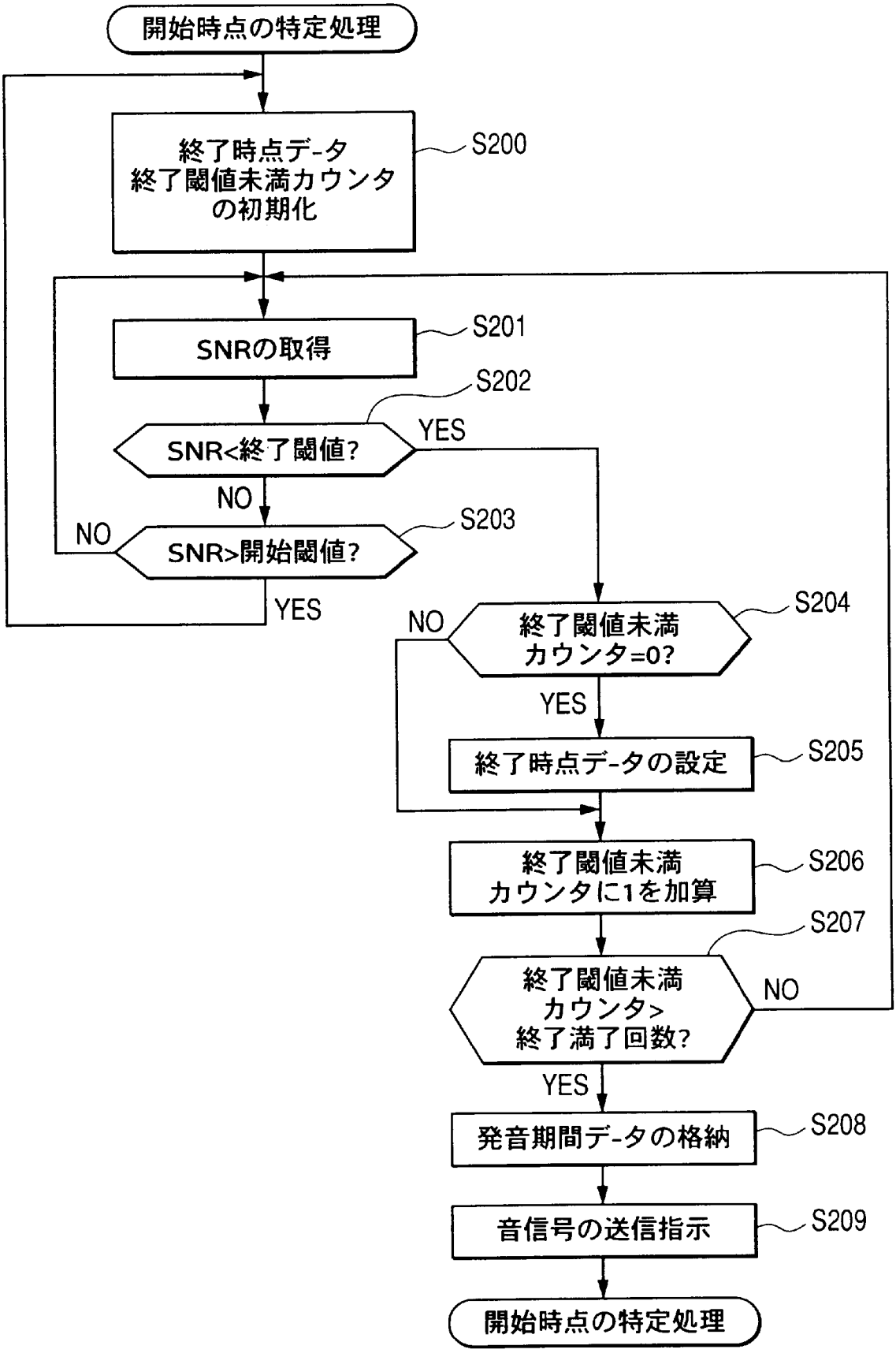
[図4]

周波数帯域番号	周波数帯域
1	200Hz～400Hz
2	400Hz～800Hz
3	800Hz～1600Hz
4	1600Hz～3200Hz
5	3200Hz～6400Hz

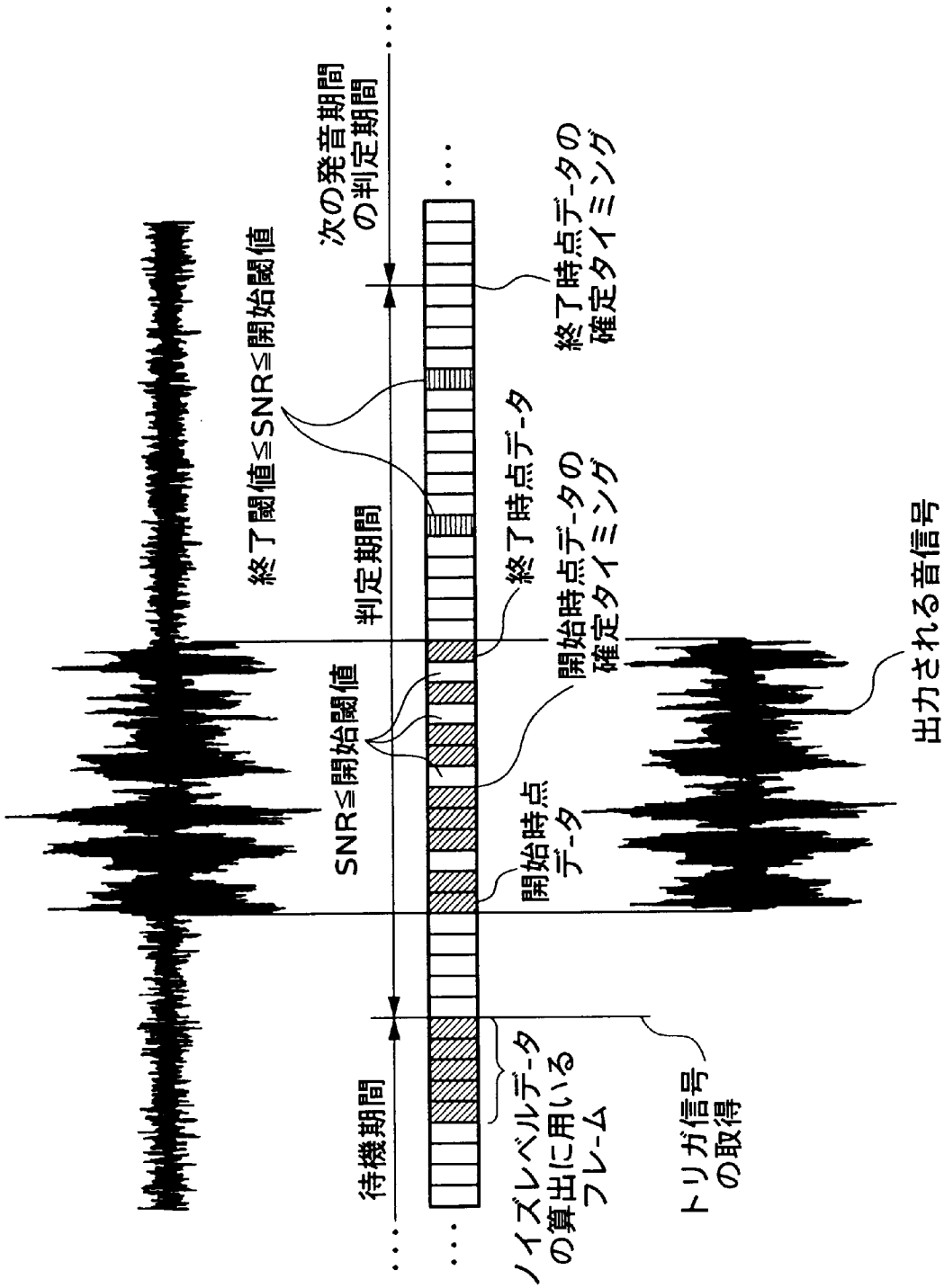
[図5]


スタート
開始時点データ
試行カウンタ
開始閾値超過カウンタ
の初期化
S100
SNRの取得
S101
試行カウンタ
に1を加算
S102
S103
SNR>開始閾値?
YES
NO
S104
NO
試行カウンタ>
開始猶予回数?
YES
NO
開始閾値超過
カウンタ=0?
S105
YES
開始時点データ
の設定
S106
開始閾値超過
カウンタに1を加算
S107
S108
開始閾値超過
カウンタ>
開始満了回数?
NO
YES
終了時点の特定処理

[図6]


開始時点の特定処理
終了時点デ-タ
終了閾値未満カウンタ
の初期化
S200
SNRの取得
S201
S202
SNR<終了閾値?
YES
NO
NO
SNR>開始閾値?
S203
YES
NO
終了閾値未満
カウンタ=0?
S204
YES
終了時点デ-タの設定
S205
終了閾値未満
カウンタに1を加算
S206
S207
終了閾値未満
カウンタ>
終了満了回数?
NO
YES
発音期間デ-タの格納
S208
音信号の送信指示
S209
開始時点の特定処理

[図7]


次の発音期間
の判定期間
終了閾値≦SNR≦開始閾値
判定期間
終了時点データ
終了時点データの
確定タイミング
SNR≦開始閾値
開始時点データの
確定タイミング
開始時点
データ
出力される音信号
待機期間
ノイズレベルデータ
の算出に用いる
フレーム
トリガ信号
の取得

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2006/312917

A. CLASSIFICATION OF SUBJECT MATTER
G10L11/02(2006.01)i, *G10L15/04*(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)
G10L11/02, 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 1996-2006
Kokai Jitsuyo Shinan Koho 1971-2006 Toroku Jitsuyo Shinan Koho 1994-2006

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 9-212195 A (Nokia Mobile Phones Ltd.), 15 August, 1997 (15.08.97), Full text; Figs. 1 to 13 & EP 784311 A1 & EP 790599 A1 & WO 1997/022116 A2 & US 5839101 A & US 5963901 A	1-18
A	JP 2004-94077 A (NEC Corp.), 25 March, 2004 (25.03.04), Full text; Figs. 1 to 5 (Family: none)	1-18
A	JP 2001-265367 A (Mitsubishi Electric Corp.), 28 September, 2001 (28.09.01), Full text; Figs. 1 to 16 (Family: none)	1-18

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

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"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)
"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
"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
"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
"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
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
01 December, 2006 (01.12.06)	12 December, 2006 (12.12.06)
Name and mailing address of the ISA/ Japanese Patent Office Facsimile No.	Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2006/312917

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 2003-524794 A (Qualcomm Inc.), 19 August, 2003 (19.08.03), Full text; Figs. 1 to 6 & EP 1159732 A1 & US 6324509 B1 & WO 2000/046790 A1	1-18
A	JP 2001-75594 A (Pioneer Electronic Corp.), 23 March, 2001 (23.03.01), Full text; Figs. 1 to 11 & EP 1081682 A2 & US 7016836 B1	1-18
A	JP 3-266899 A (Matsushita Electric Industrial Co., Ltd.), 27 November, 1991 (27.11.91), Full text; Figs. 1 to 8 (Family: none)	1-18
A	JP 7-13584 A (Matsushita Electric Industrial Co., Ltd.), 17 January, 1995 (17.01.95), Full text; Figs. 1 to 6 & US 5826230 A & US 5579431 A & US 5617508 A & WO 1996/002911 A1	1-18
A	JP 2002-73061 A (Matsushita Electric Industrial Co., Ltd.), 12 March, 2002 (12.03.02), Full text; Figs. 1 to 8 (Family: none)	1-18
A	JP 2000-163098 A (Mitsubishi Electric Corp.), 16 June, 2000 (16.06.00), Full text; Figs. 1 to 10 (Family: none)	1-18

国際調査報告

国際出願番号　ＰＣＴ／ＪＰ２００６／３１２９１７

Ａ．　発明の属する分野の分類（国際特許分類（ＩＰＣ））

Int.Cl.　G10L11/02(2006.01)i, G10L15/04(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.　G10L11/02, 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	１９７１－２００６年
日本国実用新案登録公報	１９９６－２００６年
日本国登録実用新案公報	１９９４－２００６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）

WPI

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
A	JP 9-212195 A（ノキア　モービル　フォーンズ　リミティド） 1997.08.15, 全文, 図１－13 & EP 784311 A1 & EP 790599 A1 & WO 1997/022116 A2 & US 5839101 A & US 5963901 A	１－18
A	JP 2004-94077 A（日本電気株式会社） 2004.03.25, 全文, 図１－５（ファミリーなし）	１－18
A	JP 2001-265367 A（三菱電機株式会社） 2001.09.28, 全文, 図１－16（ファミリーなし）	１－18

☑　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の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 ０１．１２．２００６	国際調査報告の発送日 １２．１２．２００６
国際調査機関の名称及びあて先 日本国特許庁（ＩＳＡ／ＪＰ） 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）　５Ｚ　８９４６ 山下　剛史 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 2003-524794 A（クゥアルコム・インコーポレイテッド） 2003.08.19，全文，図１－６ ＆ EP 1159732 A1 ＆ US 6324509 B1 ＆ WO 2000/046790 A1	１－18
A	JP 2001-75594 A（パイオニア株式会社） 2001.03.23，全文，図１－11 ＆ EP 1081682 A2 ＆ US 7016836 B1	１－18
A	JP 3-266899 A（松下電器産業株式会社） 1991.11.27，全文，第１－８図（ファミリーなし）	１－18
A	JP 7-13584 A（松下電器産業株式会社） 1995.01.17，全文，図１－６ ＆ US 5826230 A ＆ US 5579431 A ＆ US 5617508 A ＆ WO 1996/002911 A1	１－18
A	JP 2002-73061 A（松下電器産業株式会社） 2002.03.12，全文，図１－８（ファミリーなし）	１－18
A	JP 2000-163098 A（三菱電機株式会社） 2000.06.16，全文，図１－10（ファミリーなし）	１－18