

(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
2009年12月3日(03.12.2009)

PCT

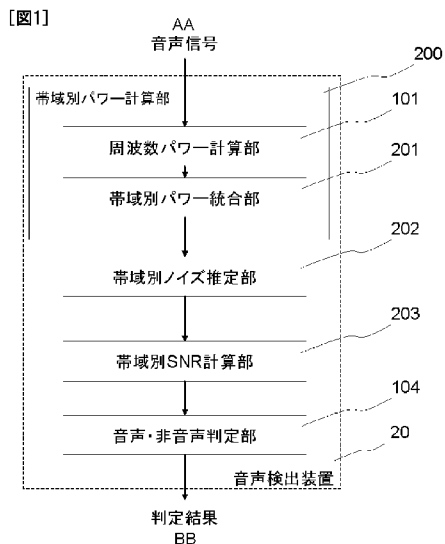
(10) 国際公開番号
WO 2009/145192 A1

- (51) 国際特許分類:
G10L 11/02 (2006.01) H04R 3/00 (2006.01)
G10L 11/00 (2006.01)
- (21) 国際出願番号: PCT/JP2009/059610
- (22) 国際出願日: 2009年5月26日(26.05.2009)
- (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 2008-139541 2008年5月28日(28.05.2008) JP
- (71) 出願人 (米国を除く全ての指定国について): 日本電気株式会社(NEC CORPORATION) [JP/JP]; 〒1088001 東京都港区芝五丁目7番1号 Tokyo (JP).
- (72) 発明者; および
- (75) 発明者/出願人 (米国についてのみ): 江森 正 (EMORI, Tadashi) [JP/JP]; 〒1088001 東京都港区芝五丁目7番1号 日本電気株式会社内 Tokyo
- (74) 代理人: 加藤 朝道 (KATO, Asamichi); 〒2220033 神奈川県横浜市港北区新横浜3丁目20番12号加藤内外特許事務所内 Kanagawa (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, AO, 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, ST, SV, 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

[続葉有]

(54) Title: VOICE DETECTION DEVICE, VOICE DETECTION METHOD, VOICE DETECTION PROGRAM, AND RECORDING MEDIUM

(54) 発明の名称: 音声検出装置、音声検出方法、音声検出プログラム及び記録媒体



AA... AUDIO SIGNAL
 200... PER-BAND POWER CALCULATION UNIT
 101... FREQUENCY POWER CALCULATION UNIT
 201... PER-BAND POWER INTEGRATION UNIT
 202... PER-BAND NOISE ESTIMATION UNIT
 203... PER-BAND SNR CALCULATION UNIT
 104... VOICE/NON-VOICE JUDGMENT UNIT
 20... VOICE DETECTION DEVICE
 BB... JUDGMENT RESULT

(57) Abstract: It is possible to improve a detection accuracy of a voice section in a conversation system allowing simultaneous utterance of more than one speakers via microphones. A voice detection device (20) includes: a per-band power calculation unit (200) which calculates for each predetermined frequency width (sub-band), a sum (sub-band power) of powers of signals inputted from respective microphones; a per-band noise estimation unit (202) which estimates a noise power for each of the sub-bands; a per-band SNR calculation unit (203) which calculates a sub-band SNR for each of the sub-bands and outputs the largest sub-band SNR as the SNR of the microphone; and a voice/non-voice judgment unit (104) which judges whether a sound is a human voice or not by using the SNR.

(57) 要約: 本発明は、複数の話者がそれぞれのマイクから同時に発声することを許容する対話システムにおける音声区間の検出精度を向上させる。音声検出装置20は、予め定めた周波数幅(サブバンド)毎に、複数のマイクからそれぞれ入力された信号のパワーの和(サブバンドパワー)を計算する帯域別パワー計算部200と、前記サブバンド毎の雑音パワーを推定する帯域別ノイズ推定部202と、前記サブバンド毎に、サブバンドSNRを計算し、一番大きなサブバンドSNRを、当該マイクのSNRとして出力する帯域別SNR計算部203と、前記SNRを用いて音声・非音声を判定する音声・非音声判定部104と、を備える(図1)。

WO 2009/145192 A1

(AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ 添付公開書類:
(AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, — 国際調査報告 (条約第 21 条(3))
GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL,
NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF,
CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD,
TG).

明 細 書

発明の名称：

音声検出装置、音声検出方法、音声検出プログラム及び記録媒体

技術分野

[0001] (関連出願についての記載)

本発明は、日本国特許出願：特願2008-139541号(2008年5月28日出願)の優先権主張に基づくものであり、同出願の全記載内容は引用をもって本書に組み込み記載されているものとする。

本発明は、音声検出装置、音声検出方法、音声検出プログラム及び記録媒体に関し、特に、複数の話者がそれぞれのマイクから同時に発声することを許容する対話システムにおける音声区間を検出するための音声検出装置、音声検出方法、音声検出プログラム及び記録媒体に関する。

背景技術

[0002] 特許文献1には、2つのマイクの出力をそれぞれ周波数帯域に分割し、これらマイクの位置に起因して変化するマイクに到達する各音響信号のパラメータ値の差を検出し、この検出差をもとに、各音響信号の周波数成分を選択して音源を分離し、目的音と目的外の音との周波数特性の違いにより識別し、目的外の音を周波数軸上で抑圧し、その出力を音源信号に合成する收音方法が開示されている。

[0003] 特許文献2には、入力時系列信号を信号分離部により分離し、分離信号に含まれる雑音成分を、複数の分離信号を用いて、雑音推定部で推定し、雑音除去部では、分離信号から推定した雑音を除去する方法が開示されている。

先行技術文献

特許文献

[0004] 特許文献1：特開2000-081900号公報

特許文献2：特開2005-308771号公報

発明の概要

発明が解決しようとする課題

[0005] なお、上記特許文献 1、2 の全開示内容はその引用をもって本書に繰込み記載する。以下の分析は、本発明によって与えられたものである。

上記した特許文献 1、2 の方法は、複数の話者の音声为重なる（クロストーク）区間の音声検出が正確にできないという問題点がある。その理由を以下に説明する。上記した特許文献 1、2 の方法は、一旦各マイクの周波数パワーの大小比較をした後、所定の帯域あるいは全帯域の周波数パワーを足し合わせることで全体のパワーを計算する。その結果、クロストーク区間のうち、全体的なパワーが大きい方の話者の声が優先されることになる。

[0006] 例えば、マイク A の前にいる話者 A が発声している最中に、マイク B の前にいる話者 B が発声した場合を考える。この場合、話者 A の音声のパワーと話者 B の音声のパワーの大小が入れ替わる時刻にて検出区間の入れ替えが起こる。このとき、話者 A については発声が終了しないうちに検出が打ち切れ、話者 B については、発声が始まってしばらく後に検出が始まるという状況になることが考えられる。更に、話者 A と話者 B の発声のタイミングによっては、マイク A、マイク B の音声が細切れに検出されることになることも考えられる。

[0007] 本発明は、上記した事情に鑑みてなされたものであって、その目的とするところは、複数の話者がそれぞれのマイクから同時に発声することを許容する対話システムにおける上記クロストーク区間の音声検出を高精度に行なうことのできる音声検出装置、音声検出方法、音声検出プログラム及び記録媒体を提供することにある。

課題を解決するための手段

[0008] 本発明の第 1 の視点によれば、予め定められた周波数幅（サブバンド）毎に、複数のマイクからそれぞれ入力された信号のパワーの和（サブバンドパワー）を計算する帯域別パワー計算部と、前記サブバンド毎の雑音パワーを推定する帯域別ノイズ推定部と、前記サブバンド毎に、サブバンド SNR（Signal to Noise Ratio）を計算し、一番大きなサブ

バンドSNRを、当該マイクのSNRとして出力する帯域別SNR計算部と、前記SNRを用いて音声・非音声を判定する音声・非音声判定部と、を備える音声検出装置が提供される。

[0009] 本発明の第2の視点によれば、複数の話者がそれぞれのマイクから同時に発声することを許容する対話システムにおける音声区間を検出するための音声検出方法であって、予め定めた周波数幅（サブバンド）毎に、複数のマイクからそれぞれ入力された信号のパワーの和（サブバンドパワー）を計算する帯域別パワー計算ステップと、前記サブバンド毎の雑音パワーを推定する帯域別ノイズ推定ステップと、前記サブバンド毎に、サブバンドSNRを計算し、一番大きなサブバンドSNRを、当該マイクのSNRとして出力する帯域別SNR計算ステップと、前記SNRを用いて音声・非音声を判定する音声・非音声判定ステップと、を含む音声検出方法が提供される。

[0010] 本発明の第3の視点によれば、複数の話者がそれぞれのマイクから同時に発声することを許容する対話システムにおける音声区間を検出するためにコンピュータに実行させる音声検出プログラムであって、予め定めた周波数幅（サブバンド）毎に、複数のマイクからそれぞれ入力された信号のパワーの和（サブバンドパワー）を計算する帯域別パワー計算処理と、前記サブバンド毎の雑音パワーを推定する帯域別ノイズ推定処理と、前記サブバンド毎に、サブバンドSNRを計算し、一番大きなサブバンドSNRを、当該マイクのSNRとして出力する帯域別SNR計算処理と、前記SNRを用いて音声・非音声を判定する音声・非音声判定処理と、を前記コンピュータに実行させる音声検出プログラム及び該プログラムを格納した記録媒体が提供される。

発明の効果

[0011] 本発明によれば、複数の話者の音声を重ねる（クロストーク）区間の音声検出を高精度に行なうことが可能となる。その理由は、複数のマイクからそれぞれ入力された信号のパワーをサブバンド毎に集計して、サブバンドSNRを計算し、一番大きなサブバンドSNRを用いて当該マイクの音声・非音

声の判定を行なうよう構成したことにある。

図面の簡単な説明

[0012] [図1]本発明の第1の実施形態に係る音声検出装置の構成を表したブロック図である。

[図2]本発明の第2の実施形態に係る音声検出装置の構成を表したブロック図である。

[図3]本発明の第3の実施形態に係る音声検出装置の構成を表したブロック図である。

[図4]本発明の第1の実施形態の音声検出装置の効果を説明するための音声検出装置の参考構成である。

[図5]クロストーク区間における音声検出の原理を説明するための図である。

発明を実施するための形態

[0013] [第1の実施形態]

続いて、本発明の第1の実施形態について図面を参照して詳細に説明する。図1は、本発明の第1の実施形態に係る音声検出装置の構成を表したブロック図である。図1を参照すると、本発明の第1の実施形態に係る音声検出装置は、帯域別パワー計算部200と、帯域別ノイズ推定部202と、帯域別SNR計算部203と、音声・非音声判定部104と、を備えた音声検出装置20が示されている。なお、上記帯域別パワー計算部200から音声・非音声判定部104までの各処理手段は、音声検出装置20を構成するコンピュータに後記する各処理を実行させ、あるいは、該コンピュータを後記各処理手段として機能させるプログラムを用いて実現することができる。

[0014] 帯域別パワー計算部200は、周波数パワー計算部101と、帯域別パワー統合部201と、を含んで構成されている。

[0015] 周波数パワー計算部101は、入力された信号を一定区間（例えば、10 msec）毎に切り出し、プリアンファシス、窓関数をかけるなどの処理を行った後、FFT（高速フーリエ変換）を行う。周波数パワー計算部101は、FFT後、一定の周波数間隔M毎のパワーを計算し、出力する。例えば

、サンプリング周波数 44.1 kHz の信号に対し、 1024 点で FFT を行った場合、約 43 Hz 間隔毎のパワーを計算することができる。これらの処理は、同時に入力された複数のマイクの信号に対しそれぞれ行われる。なお、周波数毎のパワーは、FFT 後に得られた実数部と虚数部の 2 乗和を行うことで計算できる。ここで、このような一定の周波数毎のパワーを周波数パワーと定義する。

[0016] 帯域別パワー統合部 201 は、周波数パワー計算部 101 で出力された周波数パワーを更に周波数間隔 N (但し、 $N > M$) 毎の和を計算する。ここで記述される周波数間隔 N をサブバンドと称する。また、このサブバンド毎のパワーをサブバンドパワーと呼ぶ。さらに、帯域別パワー統合部 201 は、予め定められた時間分のサブバンドパワーを保存し、その定められた時間分のサブバンドパワーの和を計算する。

[0017] サブバンドとしては、 $N > M$ となる一定の周波数間隔 N を用いることができるが、帯域に応じて和をとる幅 (周波数間隔) を変えるようにしてもよい。帯域に応じて和をとる幅 (周波数間隔) を変える例としては、音声の主要な成分を強調して表現できるメル周波数毎の間隔を挙げることができる。メル周波数毎に和を計算する場合、低周波数領域においては細かな (狭い) 間隔になり、高周波数領域については大まかな (広い) 間隔になる。なお、サブバンドパワーを保存する期間は、一定の間隔でもよいし、また、各サブバンド毎にそれぞれサブバンドパワーの保存期間を個別に設定しても良い。

[0018] 帯域別ノイズ推定部 202 は、サブバンド毎の雑音のパワーであるサブバンド雑音パワーを計算する。サブバンド雑音パワーはサブバンド毎に次の手順で計算することができる。まず、マイク毎にサブバンドパワーを比較し最もパワーの大きなマイクを選ぶ。次にマイク毎にサブバンドパワーを比較し、最小のパワーのマイクを選択し、その選ばれたマイクのサブバンドパワーを記憶する。最もパワーの大きなマイクに対応する、サブバンド雑音のパワーを前記記憶した最小のパワーとする。その他のマイクに対応する、サブバンド雑音パワーは各マイクのサブバンドパワーそのものとする。なお、その

他のマイクの雑音のパワーをそのマイクのサブバンドパワーそのものとしているのは、回り込み音声による誤検出を抑制するためである。一方、一番パワーの大きなマイクは、雑音のパワーが最小のサブバンドパワーに置き換えられるため、SNRが引き上げられることになる。

[0019] 上記帯域別ノイズ推定処理について図5を用いて説明する。サブバンド $S B_n$ において、話者A（実線）の音声パワーが最も大きいと判定され、話者Bの音声（破線）が最も小さいと判定されている場合、話者Aが用いるマイクのサブバンド雑音パワーは、話者Bのサブバンドパワーとなる。同様に、サブバンド $S B_{n+3}$ において、話者B（破線）の音声パワーが最も大きいと判定され、話者Aの音声（実線）が最も小さいと判定されている場合、話者Bが用いるマイクのサブバンド雑音パワーは、話者Aのサブバンドパワーとなる。

[0020] 帯域別SNR計算部203は、各マイクについて、サブバンド毎にサブバンドパワーをサブバンド雑音パワーで割り、サブバンド毎の信号と雑音のパワー比（SNR）を計算する。これをサブバンドSNRと呼ぶ。このようにマイク毎に計算したサブバンドSNRの中から最も大きな値のものを、そのマイクのSNRとして選択する。

[0021] 上記帯域別SNR計算処理について図5を用いて説明する。話者Aが用いるマイクのSNRのすべてのサブバンドについてサブバンドSNRが計算され、最も大きいサブバンドSNR（例えば、サブバンド $S B_n$ のサブバンドSNR）が選択される。この値が話者AのSNRとなる。同様に、話者Bが用いるマイクについても、すべてのサブバンドについてサブバンドSNRが計算され、最も大きいサブバンドSNR（例えば、サブバンド $S B_{n+3}$ のサブバンドSNR）が選択され、この値が話者BのSNRとなる。

[0022] 音声・非音声判定部104は、前記帯域別SNR計算部203で計算されたSNRを用いて、予め定められた閾値より小さい場合は非音声、予め定められた閾値より大きい場合は音声と判定する。

[0023] 上記のように帯域別SNR計算部203にて計算されるSNRは、話者毎

の声の性質や発声している内容の差で使っている周波数が違う場合があることを考慮したものとなっている（図5の話者Aと話者Bの音声パワー波形参照）。つまり、クロストーク区間であっても、図5に示すようにサブバンドレベルでピークが異なれば、それぞれの音声を検出することが可能である。従って、複数の話者の音声重なる（クロストーク）区間の音声検出の高精度化及び頑健性が確保される。

[0024] 上記本実施形態の効果をより明らかにするため、以下、サブバンドパワーの集計を行なわない音声検出装置の構成を図4を用いて説明する。ノイズ推定部102は、周波数パワー計算部101にて計算された周波数パワーに基づいて雑音のパワーを計算する。雑音のパワーは次の手順で計算される。まず、マイクごとに周波数パワーを比較し、一番パワーの大きなマイクを選ぶ。次にマイク毎に周波数パワーを比較し、最小のパワーのマイクを選択する。一番パワーの大きなマイクに対応する、雑音のパワーを、前述の最小のパワーのマイクの最小のパワーとする。その他のマイクに対応する雑音のパワーは、そのマイクの周波数パワーそのものとする。

[0025] 図4のSNR計算部103は、周波数毎に求められたパワーを全帯域に渡って足し合わせることで全帯域パワーを計算し、ノイズ推定部102において周波数毎に決定された雑音のパワーを全周波数に渡って足し合わせ全帯域雑音パワーを計算し、全帯域パワーを全帯域雑音パワーで割ることでSNRを計算する。このSNRは全マイクの信号に対してそれぞれ計算される。これは、図5の各波形全体の面積からSNRを求める処理に相当し、このとき、全体の面積が小さい話者Bの音声は検出されないことになる。

[0026] このように図4の構成では、全帯域でSNRを計算しているため、全体的なパワーが大きい方の話者の声が優先されることになる。しかし、クロストーク区間では、パワーの大小が入れ替わる時刻にて検出区間の入れ替えが起こると、先に話している話者の発声が終了しないうちに検出が打ち切れ、話者Bについては、発声が始まってしばらく後に検出が始まるといった現象が生じうる。これに対し、本実施形態の構成ではサブバンド毎に、サブバン

ドSNRを計算し、一番大きなサブバンドSNRをそのマイクのSNRとする構成を採用しているため、2以上の話者のそれぞれの周波数成分が異なるとの前提の下では、クロストーク区間における各話者の音声をそれぞれ検出できることになる。

[0027] [第2の実施形態]

続いて、各話者が用いるマイクの種類や入力音声の伝送系がそれぞれ異なる環境等への適用を考慮した本発明の第2の実施形態について説明する。複数のマイクに対しそれぞれの前に話者がいる状況において、上述した図4の構成では「入力される音声信号は話者の前にあるマイクで収録された音声のパワーが一番大きい」という仮定の基に、それぞれのマイクから得られる同じ時刻のパワーを比較し、一番大きなものを音声信号として選択している。

[0028] この仮定が成り立つのは、すべてのマイクが同じものであり、かつ各マイクと録音機器との間の接続方法が同じであることが前提とされる。一方で、これらの前提が成り立たない場合、すなわちマイクの種類が固定マイクやピンマイク等、またマイクから録音機器への伝送系が有線や無線など様々な場合も考えられる。そのような場合、マイクの種類によりその特性が大きく変わり同じ大きさの信号が入力された場合でも、マイクから得られるパワーに差異が生じる可能性がある。同様に、マイクで得られた信号が、無線、電話などの伝送系を経ることにより録音機器に到達する時刻の差異が生じる可能性も考えられる。

[0029] このような相違までを考慮に入れると、話者の前にあるマイクの音声が一番大きくなるという、図4の構成で仮定されていたことが成り立たない。さらに、伝送系の違いから遅延も生じ、「同じ時刻における信号のパワーの比較」も困難になり、音声区間の検出性能が低下することが考えられる。

[0030] 図2は、本発明の第2の実施形態に係る音声検出装置の構成を表したブロック図である。図2を参照すると、本発明における音声検出装置は、上記した第1の実施形態や図4の参考構成に示した音声検出装置20に、遅延推定部21と遅延補正部22と補正音量推定部23と音量補正部24とを追加し

た構成となっている。

- [0031] 遅延推定部 2 1 は、全マイクについて一定間隔毎に音声のパワーを計算し、パワーが急激に大きくなる時刻を測定し、一番早い時刻からの差分を計算し、遅延時間として遅延補正部 2 2 に出力する。このとき、パワーの計算は A/D 変換された区間の波形に対し、それぞれの 2 乗を足し合わせたものとして行うことができる。パワーの急激に大きくなる時刻とは、パワーが定められた閾値よりも大きくなった時刻とすることができる。
- [0032] また、上記のようにパワーそのものを閾値と比較する方法以外にも、録音開始からある一定時間を雑音であると仮定し、その区間を用いて定常雑音のパワーを推定しておき、その定常雑音のパワーと各時刻の信号のパワーの比を用いた S N R を用い、それが閾値よりも大きくなった時刻を用いてもよい。そのようにして測定された各マイクの時刻について、一番早い時刻を各マイクの時刻から引くことで、遅延時間を測定することができる。
- [0033] 遅延補正部 2 2 は、各マイクから入力された信号を、一定時間分保持し、前記遅延推定部 2 1 より出力された遅延時間だけ早めたタイミングで出力する。ここで、遅延補正部 2 2 が保持する信号量は、最低限マイク間で生じている遅延（信号の到達時間の差）以上とする。例えば、1 本目のマイクに遅延がなく、2 本目のマイクに遅延が 5 0 0 m s e c 生じている場合、遅延推定部 2 1 から遅延時間として 5 0 0 m s e c が出力される。この場合、遅延補正部 2 2 は、1 本目のマイクの信号を 5 0 0 m s e c 遅らせて出力することになる。
- [0034] より具体的には、入力された信号をサンプリング周波数 4 4 . 1 k H z、量子化ビット数 2 4 ビットで A/D 変換を行ったとき、5 0 0 m s e c 分の信号として 2 2 0 5 0 サンプルを保持しておく。この信号の保持に用いるメモリをバッファと呼ぶ。遅延補正部 2 2 は、バッファの先頭から 1 本目のマイクの信号を取り出すとともに、バッファの最後尾から 2 本のマイクの信号を取り出し、それぞれ同時に出力する。バッファ内の信号は A/D 変換された信号が入力されるとその都度新しい信号に更新される。このため、前述の

操作をし続けることで遅延のない信号を出力し続けることが可能である。

[0035] 補正音量推定部 23 は、予め定められた時間だけ各マイクの信号のパワーを計算し、計算後そのパワーを時間長で割り平均することで平均パワーを計算し、各マイクの平均パワーのうち、一番大きな値で全マイクの信号のパワーを割り、得られた値を補正係数として音量補正部 24 に出力する。ここで、補正係数の計算に用いる信号としては、すべてのマイクに均等に入力される、背景雑音のような信号を好適に用いることができる。

[0036] あるいは、一番大きなパワーの代わりに、一番小さな値や平均値など基準となるパワーを定め、これらに対する各マイクのパワーの比率を補正係数としても良い。

[0037] 音量補正部 24 は、各マイクから入力された信号に、補正音量推定部 23 より出力された補正係数を掛けて出力する。具体的には、A/D変換された信号の値に、前記補正係数を乗ずることで実現される。また、A/D変換される前のアナログ信号に対し、汎用のオーディオ機器等の増幅器を用いてもよい。この動作は、各マイクの信号に対して実施されるものとする。

[0038] 上記のように、マイクで生ずる遅延と、音量の違いを解消する機構を備えた本実施形態の音声検出装置によれば、遅延時間分のタイミングの調整と、補正係数による音量の補正が行なわれた信号が入力されるため、多種、複数マイク環境や伝送系がそれぞれ異なる環境における音声検出の精度を上げることが可能である。

[0039] 特に、上記した第 1 の実施形態の音声検出装置に適用すれば、クロストーク区間における音声検出精度をより向上させることができる。もちろん、図 4 に示した音声検出装置に適用しても、多種、複数マイク環境や伝送系がそれぞれ異なる環境における音声検出の精度を上げることが可能である。

[0040] [第 3 の実施形態]

続いて、上記本発明の第 2 の実施形態に改良を加えた本発明の第 3 の実施形態について説明する。

[0041] 図 3 は、本発明の第 3 の実施形態に係る音声検出装置の構成を表したブロ

ック図である。図3を参照すると、本発明における音声検出装置は、上記した第2の実施形態に、突発音発生部25を追加した構成となっている。

[0042] 突発音発生部25は、所定の起動手手段（スイッチ）により動作し、大きな音（突発音）を出力する。突発音としては、全周波数にわたり、かつ急激にパワーの大きくなる音が望ましい。

[0043] 突発音発生部25より出力された突発音により、前記遅延推定部21又は補正音量推定部23、あるいはその両方を動作させることで、遅延時間及び補正係数の測定精度を向上させることが可能である。例えば、多種・複数マイクがセットされた部屋で、しばらく静かにしておき、突発音発生部25を作動させることで遅延時間及び補正係数がそれぞれ正確に計算される。

[0044] 以上、本発明の好適な実施形態を説明したが、本発明は、上記した実施形態に限定されるものではなく、本発明の基本的技術的思想を逸脱しない範囲で、更なる変形・置換・調整を加えることができる。例えば、遅延が生じない環境では、上記した第2、第3の実施形態の遅延推定部21と遅延補正部22とを省略することができる。同様に、マイク間で音量の差が生じない環境では、上記した第2の実施形態の補正音量推定部23と音量補正部24とを省略することができる。

[0045] また、上記した第1の実施形態では、周波数パワー計算部101と、帯域別パワー統合部201との構成で、帯域別パワー（サブバンドパワー）を計算するものとして説明したが、周波数パワー計算部101及び帯域別パワー統合部201における各処理を1つの処理ブロックで実行する構成も採用可能である。

[0046] また、上記した実施形態で示した信号パワーやSNRの計算式は、それぞれの説明に好適な例を示したものであり、当業者が採用できる各種の計算方法を採用できることはいうまでもない。

産業上の利用可能性

[0047] 本発明によれば、音声検出を行う音声検出装置や、音声検出装置をコンピュータに実現するためのプログラムといった用途に適用できる。

本発明の全開示（請求の範囲を含む）の枠内において、さらにその基本的技術思想に基づいて、実施例ないし実施例の変更・調整が可能である。また、本発明の請求の範囲の枠内において種々の開示要素の多様な組み合わせないし選択が可能である。すなわち、本発明は、請求の範囲を含む全開示、技術的思想にしたがって当業者であればなし得るであろう各種変形、修正を含むことは勿論である。

符号の説明

- [0048] 10、20 音声検出装置
- 21 遅延推定部
 - 22 遅延補正部
 - 23 補正音量推定部
 - 24 音量補正部
 - 25 突発音発生部
- 101 周波数パワー計算部
 - 102 ノイズ推定部
 - 103 SNR計算部
 - 104 音声・非音声判定部
- 200 帯域別パワー計算部
 - 201 帯域別パワー統合部
 - 202 帯域別ノイズ推定部
 - 203 帯域別SNR計算部

請求の範囲

- [請求項1] 予め定めた周波数幅（サブバンド）毎に、複数のマイクからそれぞれ入力された信号のパワーの和（サブバンドパワー）を計算する帯域別パワー計算部と、
- 前記サブバンド毎の雑音パワーを推定する帯域別ノイズ推定部と、
- 前記サブバンド毎に、サブバンドSNRを計算し、一番大きなサブバンドSNRを、当該マイクのSNRとして出力する帯域別SNR計算部と、
- 前記SNRを用いて音声・非音声を判定する音声・非音声判定部と、
- を備える音声検出装置。
- [請求項2] 前記帯域別ノイズ推定部は、前記サブバンドパワーをマイク毎に比較し、サブバンドパワーが大きいマイクと小さいマイクをそれぞれ一つずつ選択し、前記サブバンドパワーが大きい方のマイクの当該サブバンドに対応するサブバンド雑音パワーを、前記サブバンドパワーが小さいマイクのサブバンドパワーとする請求項1に記載の音声検出装置。
- [請求項3] 前記帯域別ノイズ推定部は、その他のマイクのサブバンド雑音パワーを当該マイクのサブバンドパワーとする請求項2に記載の音声検出装置。
- [請求項4] 前記サブバンドは、低周波数域では狭く、高周波数域では広い間隔となるよう設定されている請求項1乃至3いずれかーに記載の音声検出装置。
- [請求項5] 更に、前記複数のマイクから入力された信号の遅延を補正する遅延補正部を備える、請求項1乃至4いずれかーに記載の音声検出装置。
- [請求項6] 更に、前記複数のマイクから入力された信号の音量を補正する音量補正部を備える、請求項1乃至5いずれかーに記載の音声検出装置。
- [請求項7] 更に、前記各マイクそれぞれの信号のパワーが大きく変化する時刻を測定しておき、それら時刻の差分を遅延時間として前記遅延補正部

に出力する遅延時間測定部を備える、
請求項 5 又は 6 に記載の音声検出装置。

[請求項 8] 更に、前記各マイクそれぞれの信号のパワーの比を計算し前記音量を補正する際の補正係数を前記音量補正部に出力する補正音量推定部を備える、
請求項 6 又は 7 に記載の音声検出装置。

[請求項 9] 更に、短時間に突発的な音を出力する突発音発生部を備える、
請求項 7 又は 8 に記載の音声検出装置。

[請求項 10] 前記帯域別パワー計算部は、所定の周波数幅（サブバンド）毎に、所定の時間範囲の周波数毎のパワーの和（サブバンドパワー）を計算する請求項 1 乃至 9 いずれかーに記載の音声検出装置。

[請求項 11] 複数の話者がそれぞれのマイクから同時に発声することを許容する対話システムにおける音声区間を検出するための音声検出方法であって、

予め定めた周波数幅（サブバンド）毎に、複数のマイクからそれぞれ入力された信号のパワーの和（サブバンドパワー）を計算する帯域別パワー計算ステップと、

前記サブバンド毎の雑音パワーを推定する帯域別ノイズ推定ステップと、

前記サブバンド毎に、サブバンド SNR を計算し、一番大きなサブバンド SNR を、当該マイクの SNR として出力する帯域別 SNR 計算ステップと、

前記 SNR を用いて音声・非音声を判定する音声・非音声判定ステップと、を含むことを特徴とする音声検出方法。

[請求項 12] 帯域別ノイズ推定ステップにおいて、前記サブバンドパワーをマイク毎に比較し、サブバンドパワーが大きいマイクと小さいマイクをそれぞれ一つずつ選択し、前記サブバンドパワーが大きい方のマイクの当該サブバンドに対応するサブバンド雑音パワーを、前記サブバンド

パワーが小さいマイクのサブバンドパワーとする請求項 1 1 に記載の音声検出方法。

[請求項13] 前記帯域別ノイズ推定ステップにおいて、その他のマイクのサブバンド雑音パワーを当該マイクのサブバンドパワーとする請求項 1 2 に記載の音声検出方法。

[請求項14] 前記サブバンドは、低周波数域では狭く、高周波数域では広い間隔となるよう設定されている請求項 1 1 乃至 1 3 いずれかーに記載の音声検出方法。

[請求項15] 更に、前記複数のマイクから入力された信号の遅延を補正する遅延補正ステップを含む請求項 1 1 乃至 1 4 いずれかーに記載の音声検出方法。

[請求項16] 更に、前記複数のマイクから入力された信号の音量を補正する音量補正ステップを含む請求項 1 1 乃至 1 5 いずれかーに記載の音声検出方法。

[請求項17] 更に、前記各マイクそれぞれの信号のパワーが大きく変化する時刻を測定しておき、それら時刻の差分を遅延時間として出力する遅延時間測定ステップを含み、

前記遅延補正ステップにおいて、前記遅延時間分の補正を行なう請求項 1 5 又は 1 6 に記載の音声検出方法。

[請求項18] 更に、前記各マイクそれぞれの信号のパワーの比を計算し、前記音量を補正する際の補正係数を出力する補正音量推定ステップを含み、

前記音量補正ステップにおいて、前記補正係数を用いた補正を行なう請求項 1 6 又は 1 7 に記載の音声検出方法。

[請求項19] 更に、短時間に突発的な音を出力する突発音発生部からの出力信号に基づいて、前記各マイクそれぞれの信号の遅延時間又はパワーの比を計算する請求項 1 7 又は 1 8 に記載の音声検出方法。

[請求項20] 前記帯域別パワー計算ステップにおいて、所定の周波数幅（サブバンド）毎に、所定の時間範囲の周波数毎のパワーの和（サブバンドパ

ワー) を計算する請求項 11 乃至 19 いずれかーに記載の音声検出方法。

[請求項21]

複数の話者がそれぞれのマイクから同時に発声することを許容する対話システムにおける音声区間を検出するためにコンピュータに実行させる音声検出プログラムであって、

予め定めた周波数幅(サブバンド)毎に、複数のマイクからそれぞれ入力された信号のパワーの和(サブバンドパワー)を計算する帯域別パワー計算処理と、

前記サブバンド毎の雑音パワーを推定する帯域別ノイズ推定処理と、

、

前記サブバンド毎に、サブバンドSNRを計算し、一番大きなサブバンドSNRを、当該マイクのSNRとして出力する帯域別SNR計算処理と、

前記SNRを用いて音声・非音声を判定する音声・非音声判定処理と、を前記コンピュータに実行させる音声検出プログラム。

[請求項22]

前記帯域別ノイズ推定処理において、前記サブバンドパワーをマイク毎に比較し、サブバンドパワーが大きいマイクと小さいマイクをそれぞれ一つずつ選択し、前記サブバンドパワーが大きい方のマイクの当該サブバンドに対応するサブバンド雑音パワーを、前記サブバンドパワーが小さいマイクのサブバンドパワーとする処理を実行する請求項 21 に記載の音声検出プログラム。

[請求項23]

前記帯域別ノイズ推定処理において、その他のマイクのサブバンド雑音パワーを当該マイクのサブバンドパワーとする請求項 22 に記載の音声検出プログラム。

[請求項24]

前記サブバンドは、低周波数域では狭く、高周波数域では広い間隔となるよう設定されている請求項 21 乃至 23 いずれかーに記載の音声検出プログラム。

[請求項25]

更に、前記複数のマイクから入力された信号の遅延を補正する遅延

補正処理を実行させる請求項 2 1 乃至 2 4 いずれかーに記載の音声検出プログラム。

[請求項26] 更に、前記複数のマイクから入力された信号の音量を補正する音量補正処理を実行させる請求項 2 1 乃至 2 5 いずれかーに記載の音声検出プログラム。

[請求項27] 前記各マイクそれぞれの信号のパワーが大きく変化する時刻を測定しておき、それら時刻の差分を遅延時間として出力する遅延時間測定処理を実行させ、

前記遅延補正処理において、前記遅延時間分の補正を行なう請求項 2 5 又は 2 6 に記載の音声検出プログラム。

[請求項28] 更に、前記各マイクそれぞれの信号のパワーの比を計算し、前記音量を補正する際の補正係数を出力する補正音量推定処理を実行させ、

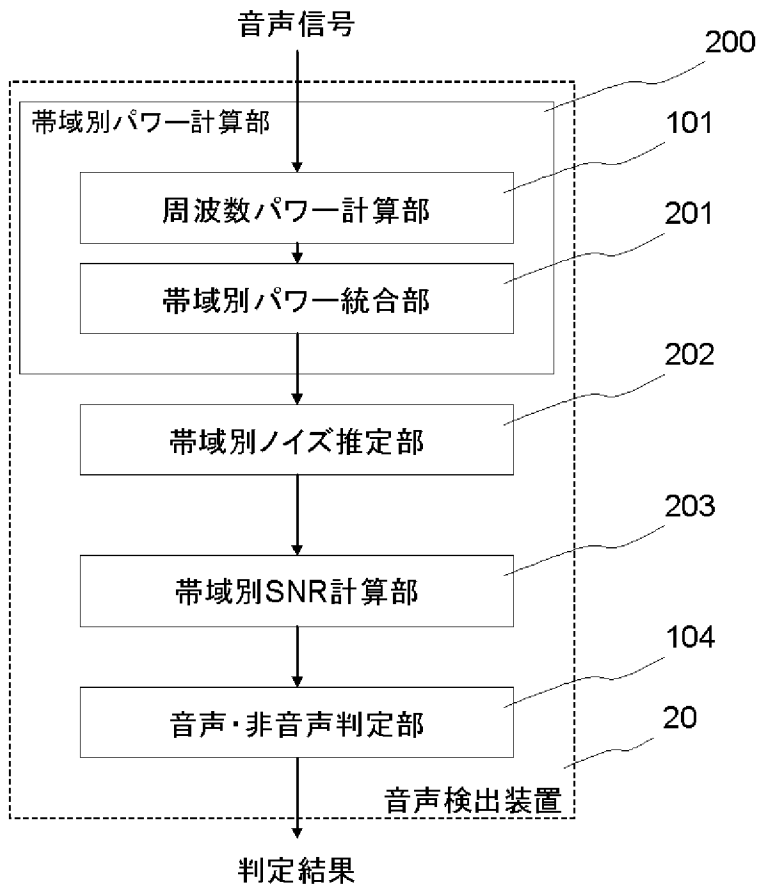
前記音量補正処理において、前記補正係数を用いた補正を行なう請求項 2 6 又は 2 7 に記載の音声検出プログラム。

[請求項29] 更に、短時間に突発的な音を出力する突発音発生部を作動させ、該突発音発生部からの出力信号に基づいて、前記各マイクそれぞれの信号の遅延時間又はパワーの比を計算する請求項 2 7 又は 2 8 に記載の音声検出プログラム。

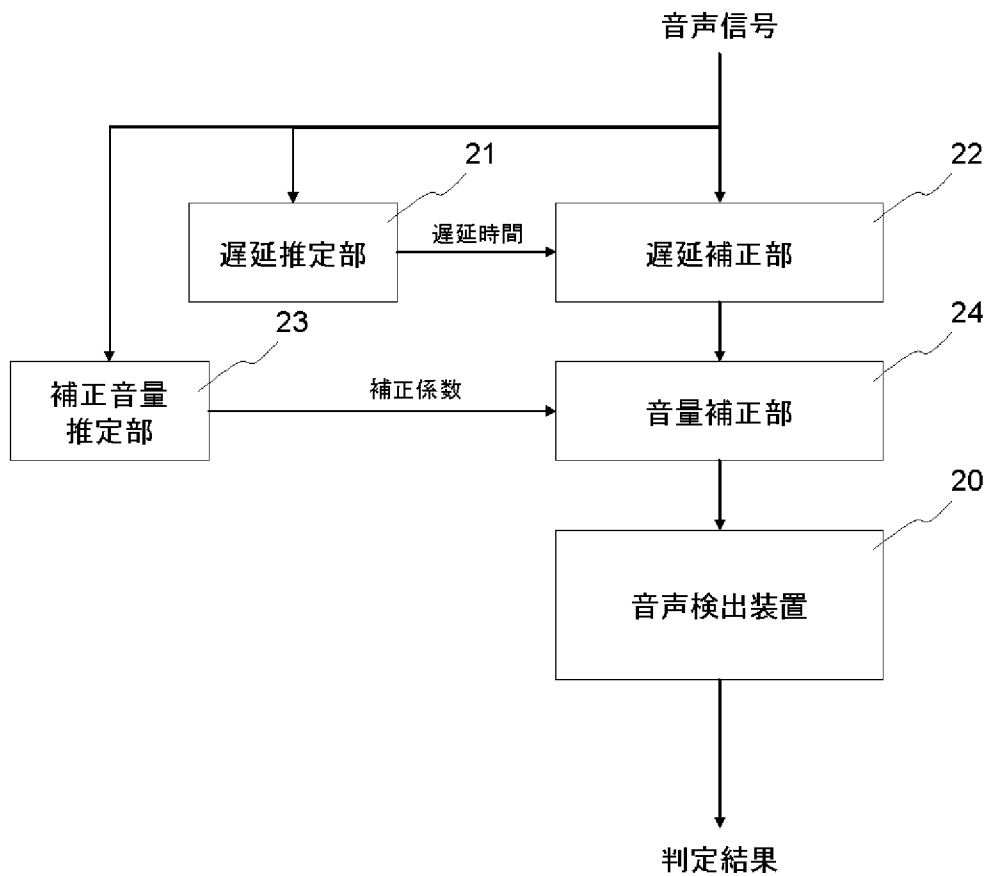
[請求項30] 前記帯域別パワー計算処理において、所定の周波数幅（サブバンド）毎に、所定の時間範囲の周波数毎のパワーの和（サブバンドパワー）を計算する請求項 2 1 乃至 2 9 いずれかーに記載の音声検出プログラム。

[請求項31] 請求項 2 1 乃至 3 0 いずれかーに記載の音声検出プログラムを格納した記録媒体。

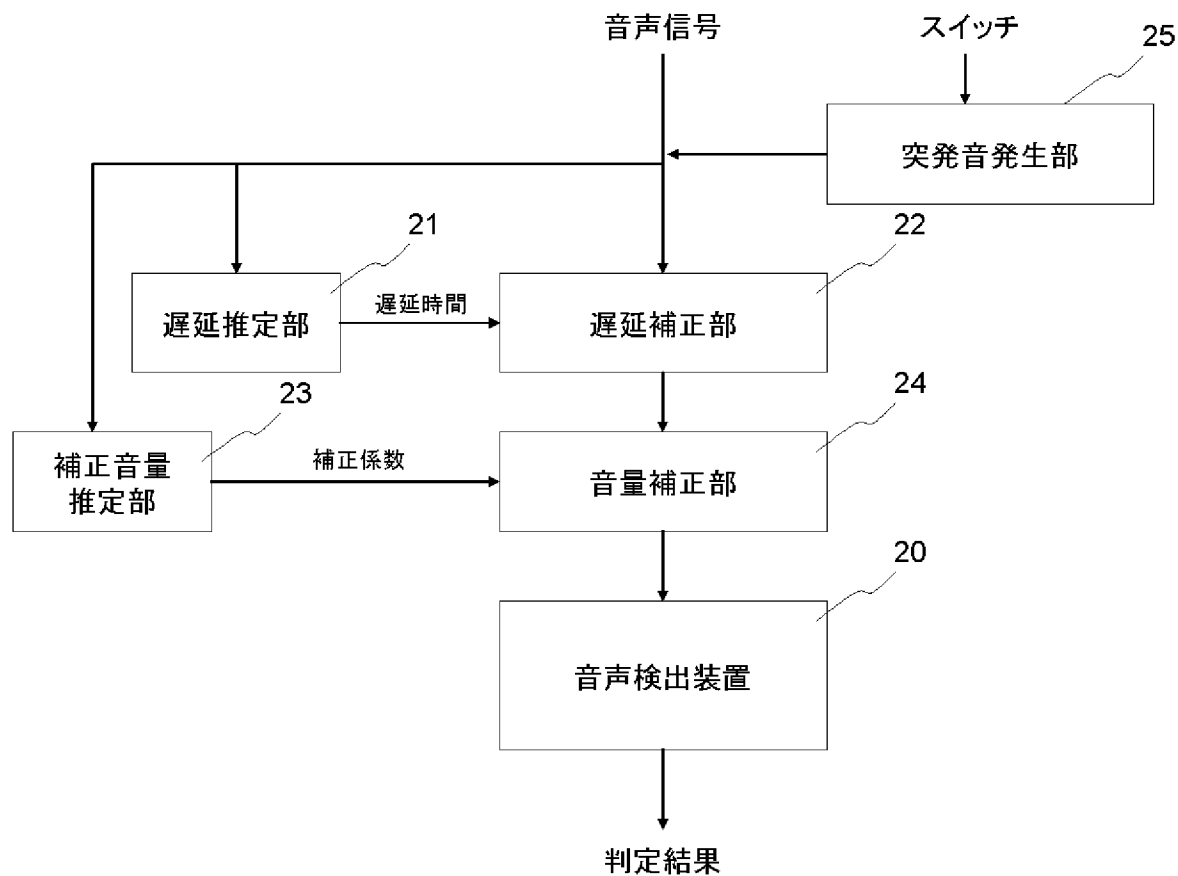
[図1]



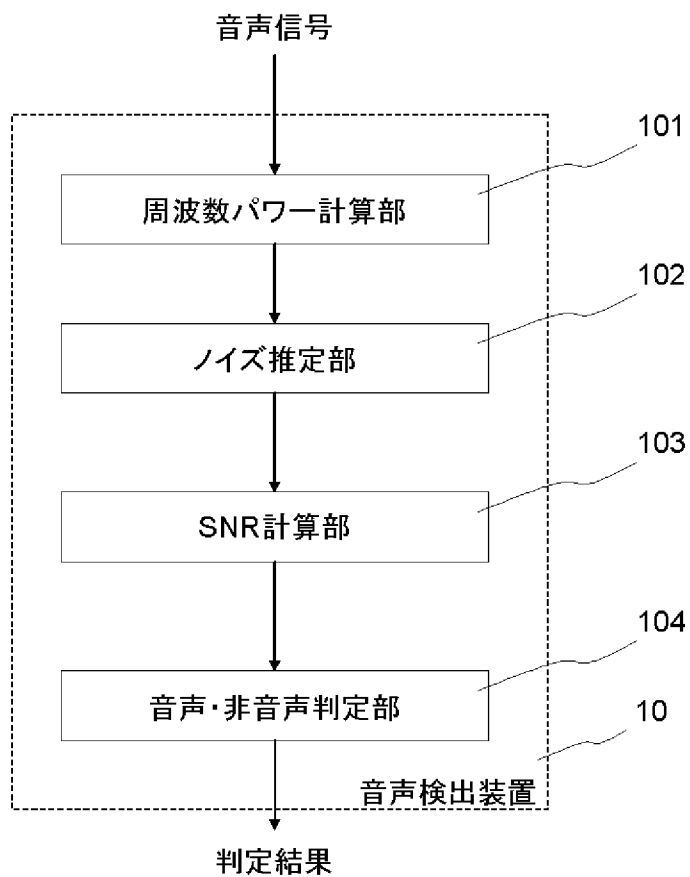
[図2]



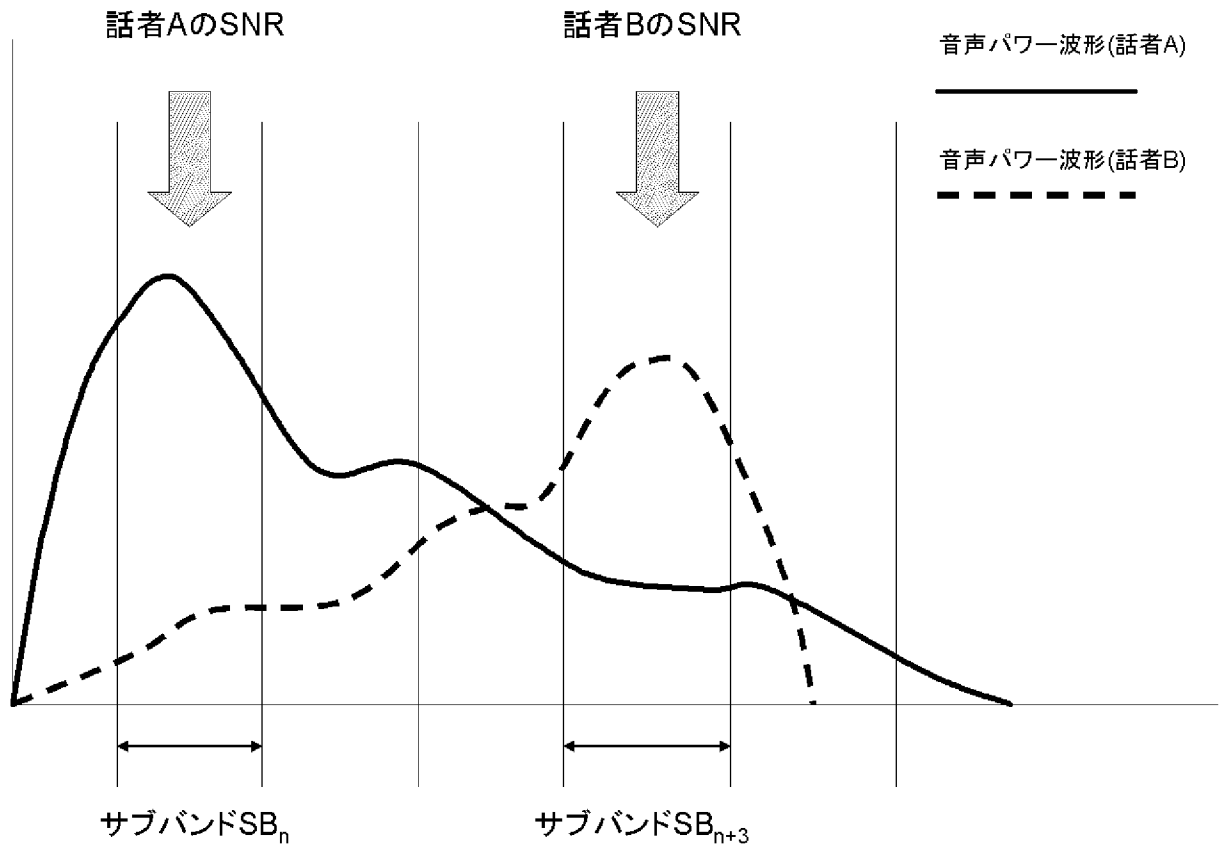
[図3]



[図4]



[図5]



INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2009/059610

A. CLASSIFICATION OF SUBJECT MATTER
G10L11/02(2006.01) i, G10L11/00(2006.01) i, H04R3/00(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)
G10L11/02, G10L11/00, H04R3/00

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	1996-2009
Kokai Jitsuyo Shinan Koho	1971-2009	Toroku Jitsuyo Shinan Koho	1994-2009

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	JP 3588030 B2 (Mitsubishi Electric Corp.), 20 August, 2004 (20.08.04), Par. No. [0046]	1, 4-11, 14-21, 24-31
Y	JP 3163109 B2 (Oki Electric Industry Co., Ltd.), 23 February, 2001 (23.02.01), Fig. 1	1, 4, 10, 11, 14, 20, 21, 24, 30, 31
Y	JP 2007-68125 A (NEC Corp.), 15 March, 2007 (15.03.07), Full text; all drawings	5-8, 15-18, 25-28
Y	JP 2002-502193 A (Telefonaktiebolaget LM Ericsson (publ)), 22 January, 2002 (22.01.02), Full text; all drawings	9, 19, 29

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 30 June, 2009 (30.06.09)	Date of mailing of the international search report 07 July, 2009 (07.07.09)
---	--

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

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2009/059610

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 9-212195 A (Nokia Mobile Phones Ltd.), 15 August, 1997 (15.08.97), Full text; all drawings	1-31
A	JP 3218681 B2 (Sony Corp.), 10 August, 2001 (10.08.01), Full text; all drawings	1-31

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/JP2009/059610

JP 3588030 B2	2004.08.20	(Family: none)
JP 3163109 B2	2001.02.23	(Family: none)
JP 2007-68125 A	2007.03.15	CN 1925693 A EP 1804549 A2 KR 2007-0026235 A US 2007/053455 A1
JP 2002-502193 A	2002.01.22	AT 349849 T AU 2646699 A BR 9908204 A CN 1289508 A EP 1051835 B1 US 6549627 B1 WO 1999/039497 A1
JP 9-212195 A	1997.08.15	AU 1067797 A AU 1067897 A DE 69614989 T2 DE 69630580 T2 EP 784311 B1 EP 790599 B1 FI 100840 B US 5839101 A WO 1997/022116 A2 WO 1997/022117 A1
JP 3218681 B2	2001.08.10	(Family: none)

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, G10L11/00(2006.01)i, H04R3/00(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. G10L11/02, G10L11/00, H04R3/00

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 1971-2009年
 日本国実用新案登録公報 1996-2009年
 日本国登録実用新案公報 1994-2009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)
 CiNii

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 請求項の番号
Y	JP 3588030 B2(三菱電機株式会社) 2004.08.20, 段落【0046】	1, 4-11, 14-21 , 24-31
Y	JP 3163109 B2(沖電気工業株式会社) 2001.02.23, 【図1】	1, 4, 10, 11, 14 , 20, 21, 24, 30 , 31
Y	JP 2007-68125 A(日本電気株式会社) 2007.03.15, 全文, 全図	5-8, 15-18, 25-28

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。

<p>* 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</p>	<p>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</p>
--	---

国際調査を完了した日 30.06.2009	国際調査報告の発送日 07.07.2009
--------------------------	--------------------------

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) 涌井 智則	5 Z	9 5 6 8
	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 請求項の番号
Y	JP 2002-502193 A(テレフォンアクチーボラゲット エル エム エリクソン (パブル)) 2002. 01. 22, 全文, 全図	9, 19, 29
A	JP 9-212195 A(ノキア モービル フォーンズ リミテッド) 1997. 08. 15, 全文, 全図	1-31
A	JP 3218681 B2(ソニー株式会社) 2001. 08. 10, 全文, 全図	1-31

JP 3588030 B2	2004. 08. 20	ファミリーなし
JP 3163109 B2	2001. 02. 23	ファミリーなし
JP 2007-68125 A	2007. 03. 15	CN 1925693 A EP 1804549 A2 KR 2007-0026235 A US 2007/053455 A1
JP 2002-502193 A	2002. 01. 22	AT 349849 T AU 2646699 A BR 9908204 A CN 1289508 A EP 1051835 B1 US 6549627 B1 WO 1999/039497 A1
JP 9-212195 A	1997. 08. 15	AU 1067797 A AU 1067897 A DE 69614989 T2 DE 69630580 T2 EP 784311 B1 EP 790599 B1 FI 100840 B US 5839101 A WO 1997/022116 A2 WO 1997/022117 A1
JP 3218681 B2	2001. 08. 10	ファミリーなし