

(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
2018年12月27日(27.12.2018)



(10) 国際公開番号
WO 2018/235380 A1

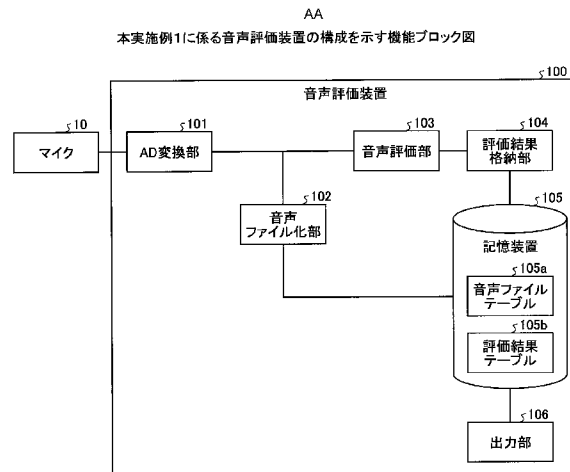
- (51) 国際特許分類:
G10L 25/60 (2013.01) *G10L 25/90* (2013.01)
- (21) 国際出願番号: PCT/JP2018/013867
- (22) 国際出願日: 2018年3月30日(30.03.2018)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:
特願 2017-123588 2017年6月23日(23.06.2017) JP
- (71) 出願人: 富士通株式会社 (FUJITSU LIMITED)
[JP/JP]; 〒2118588 神奈川県川崎市中原区上小田中4丁目1番1号 Kanagawa (JP).
- (72) 発明者: 中山 紗友梨 (NAKAYAMA, Sayuri);
〒2118588 神奈川県川崎市中原区上小田中4

丁目1番1号 富士通株式会社内 Kanagawa (JP). 外川 太郎 (TOGAWA, Taro); 〒2118588 神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内 Kanagawa (JP). 大谷 猛 (OTANI, Takeshi); 〒2118588 神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内 Kanagawa (JP).

- (74) 代理人: 向山 直樹 (MUKOUYAMA Naoki);
〒2118588 神奈川県川崎市中原区上小田中4丁目1番1号 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, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,

(54) Title: VOICE EVALUATION PROGRAM, VOICE EVALUATION METHOD, AND VOICE EVALUATION DEVICE

(54) 発明の名称: 音声評価プログラム、音声評価方法および音声評価装置



10 Microphone
 100 Voice evaluation device
 101 A-D conversion unit
 102 Voice filing unit
 103 Voice evaluation unit
 104 Evaluation result storage unit
 105 Storage device
 105a Voice file table
 105b Evaluation result table
 106 Output unit
 AA Functional block diagram indicating configuration of voice evaluation device relating to embodiment 1

(57) Abstract: [Problem] The present invention addresses the problem of accurately evaluating a voice. [Solution] A voice evaluation device 100 analyzes voice signals, and detects pitch frequencies. On the basis of the detection frequency distribution of the pitch frequencies thus detected, the voice evaluation device 100 selects a region to be evaluated from among the pitch frequencies thus detected. On the basis of the detection frequency distribution and the selected region to be evaluated, the voice evaluation device 100 evaluates a voice.

WO 2018/235380 A1

HN, HR, HU, ID, IL, IN, IR, IS, JO, KE, KG, KH,
KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY,
MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ,
NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT,
QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,
SM, ST, SV, SY, TH, 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, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

添付公開書類 :

- 一 国際調査報告 (条約第21条(3))

(57) 要約: 【課題】 音声を正確に評価すること。 【解決手段】 音声評価装置100は、音声信号を解析してピッチ周波数を検出する。音声評価装置100は、検出したピッチ周波数の検出頻度の分布に基づいて、検出したピッチ周波数のうち評価対象とする評価対象領域を選択する。音声評価装置100は、検出頻度の分布と選択した評価対象領域に基づいて、音声を評価する。

明 細 書

発明の名称：

音声評価プログラム、音声評価方法および音声評価装置

技術分野

[0001] 本発明は、音声評価プログラム等に関する。

背景技術

[0002] 近年、多くの企業が顧客満足度を重視する傾向にあり、対応者と顧客との会話における、対応者の印象や顧客の感情を把握したいというニーズが高まっている。対応者の印象や、顧客の感情は、音声に現れることが多い。

[0003] たとえば、従来技術には、通話中の音声のピッチ周波数の上下幅に基づいて、音声の印象を判定するものがある。図15は、従来技術を説明するための図である。図15のグラフ5の横軸は時間を示す軸であり、縦軸は周波数を示す軸である。

[0004] 従来技術では、ピッチ周波数の上下幅が大きいものを明瞭（明るい）と判定し、ピッチ周波数の上下幅が小さいものを不明瞭（暗い）と判定する。ピッチ周波数の上下幅は、ある期間中のピッチ周波数の最大値と最小値との差である。

[0005] たとえば、従来技術は、グラフ5の期間T1において、上下幅5aが大きいため、音声の印象が明瞭であると判定する。従来技術は、グラフ5の期間T2において、上下幅5bが小さいため、音声の印象が不明瞭であると判定する。

先行技術文献

特許文献

[0006] 特許文献1：特開2015-87557号公報

特許文献2：特開2006-267465号公報

発明の概要

発明が解決しようとする課題

[0007] しかしながら、上述した従来技術では、音声を正確に評価することができないという問題がある。

[0008] ピッチ周波数を検出する処理では、特性上、半ピッチや倍ピッチを誤って算出する場合がある。このため、従来技術のように、単純に、ピッチ周波数の上下幅の大小により、音声の明瞭・不明瞭を判定すると、音声を正確に評価できない場合がある。

[0009] 図16は、正確なピッチ周波数を算出した場合のグラフを示す図である。図16のグラフ6の横軸は時間を示す軸であり、縦軸は周波数を示す軸である。グラフ6は、正確なピッチ周波数を算出した場合を示すものである。グラフ6では、ピッチ周波数の上下幅6aが小さいため、音声不明瞭であると判定できる。

[0010] 図17は、半ピッチ／倍ピッチを算出した場合のグラフを示す図である。図17のグラフ7の横軸は時間を示す軸であり、縦軸は周波数を示す軸である。グラフ7は、誤って、半ピッチ／倍ピッチを算出した場合を示すものである。グラフ7では、ピッチ周波数の上下幅7aが大きくなってしまい、実際には、音声不明瞭であるにもかかわらず、音声明瞭であると判定してしまう。

[0011] 1つの側面では、本発明は、音声を正確に評価することができる音声評価プログラム、音声評価方法および音声評価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

[0012] 第1の案では、コンピュータに下記の処理を実行させる。コンピュータは、音声信号を解析してピッチ周波数を検出する。コンピュータは、検出したピッチ周波数の検出頻度の分布に基づいて、検出したピッチ周波数のうち評価対象とする評価対象領域を選択する。コンピュータは、検出頻度の分布と選択した評価対象領域に基づいて、音声を評価する。

発明の効果

[0013] 音声を正確に評価することができる。

図面の簡単な説明

[0014] [図1]図 1 は、本実施例 1 に係る音声評価装置の構成を示す機能ブロック図である。

[図2]図 2 は、本実施例 1 に係る音声評価部の構成を示す機能ブロック図である。

[図3]図 3 は、ヒストグラムの一例を示す図である。

[図4]図 4 は、本実施例 1 に係る音声評価部の処理手順を示すフローチャートである。

[図5]図 5 は、音声評価装置のその他の処理を説明するための図である。

[図6]図 6 は、本実施例 2 に係る音声評価システムの構成を示す図である。

[図7]図 7 は、本実施例 2 に係る音声評価装置の構成を示す機能ブロック図である。

[図8]図 8 は、本実施例 2 に係る音声評価部の構成を示す機能ブロック図である。

[図9]図 9 は、本実施例 2 に係る平常時分布テーブルのデータ構造の一例を示す図である。

[図10]図 10 は、本実施例 3 に係る音声評価システムの構成を示す図である。

[図11]図 11 は、本実施例 3 に係る収録機器の構成を示す機能ブロック図である。

[図12]図 12 は、本実施例 3 に係る音声評価装置の構成を示す機能ブロック図である。

[図13]図 13 は、本実施例 3 に係る音声評価部の構成を示す機能ブロック図である。

[図14]図 14 は、音声評価装置と同様の機能を実現するコンピュータのハードウェア構成の一例を示す図である。

[図15]図 15 は、従来技術を説明するための図である。

[図16]図 16 は、正確なピッチ周波数を算出した場合のグラフを示す図であ

る。

[図17]図17は、半ピッチ／倍ピッチを算出した場合のグラフを示す図である。

発明を実施するための形態

[0015] 以下に、本願の開示する音声評価プログラム、音声評価方法および音声評価装置の実施例を図面に基づいて詳細に説明する。なお、この実施例によりこの発明が限定されるものではない。

実施例 1

[0016] 図1は、本実施例1に係る音声評価装置の構成を示す機能ブロック図である。図1に示すように、この音声評価装置100は、話者（図示略）の音声を集音するマイク10に接続される。マイク10は、集音した音声の信号を、音声評価装置100に出力する。以下の説明では、マイク10が集音した音声の信号を「音声信号」と表記する。

[0017] 音声評価装置100は、AD（Analog-to-Digital）変換部101、音声ファイル化部102、音声評価部103、評価結果格納部104、記憶装置105、出力部106を有する。

[0018] AD変換部101は、マイク10から音声信号を受信し、AD変換を実行する処理部である。具体的には、AD変換部101は、音声信号（アナログ信号）を、音声信号（デジタル信号）に変換する。AD変換部101は、音声信号（デジタル信号）を、音声ファイル化部102、音声評価部103に出力する。以下の説明では、AD変換部101から出力される音声信号（デジタル信号）を単に音声信号と表記する。

[0019] 音声ファイル化部102は、音声信号を所定の音声ファイルフォーマットにより、音声ファイルに変換する処理部である。たとえば、音声ファイルは、各時刻と、音声信号の強さとをそれぞれ対応づけた情報を含む。音声ファイル化部102は、音声ファイルを、記憶部105の音声ファイルテーブル

105aに格納する。以下の説明では、便宜上、音声ファイルに含まれる時刻と音声信号の強さとの関係の情報を単に、音声信号と記載する。

[0020] なお、音声ファイル化部102は、入力装置（図示略）から話者情報を取得し、話者情報を音声ファイルに添付する。たとえば、話者情報は、話者を一意に識別する情報である。

[0021] 音声評価部103は、音声信号を基にして、話者の音声の印象を評価する処理部である。たとえば、話者の音声の印象の評価結果は「印象が良い」、「普通」、「印象が悪い」のいずれかとなる。音声評価部103は、話者情報と、評価結果の情報とを、評価結果格納部104に出力する。

[0022] 評価結果格納部104は、話者情報と、評価結果とを対応づけて、記憶装置105の評価結果テーブル105bに格納する処理部である。

[0023] 記憶装置105は、音声ファイルテーブル105aおよび評価結果テーブル105bを記憶する。記憶装置105は、RAM (Random Access Memory)、ROM (Read Only Memory)、フラッシュメモリ (Flash Memory) などの半導体メモリ素子や、HDD (Hard Disk Drive) などの記憶装置に対応する。

[0024] 音声ファイルテーブル105aは、音声ファイル化部102から出力される音声ファイルを格納するテーブルである。評価結果テーブル105bは、評価結果格納部104により格納される話者情報と、評価結果とを対応づけたテーブルである。

[0025] 出力部106は、記憶装置105に格納された評価結果テーブル105bを、表示装置に出力することで、評価結果を表示させる処理部である。また、出力部106は、音声ファイルテーブル105aに格納された音声ファイルを他の外部装置に出力してもよい。

[0026] ここで、図1に示したAD変換部101、音声ファイル化部102、音声評価部103、評価結果格納部104、出力部106は、所定の制御部（図示略）に対応する。たとえば、制御部は、CPU (Central Processing Unit) やMPU (Micro Processing Unit) などによって実現できる。また、

制御部は、ASIC (Application Specific Integrated Circuit) やFPGA (Field Programmable Gate Array) などのハードワイヤードロジックによっても実現できる。

[0027] 続いて、図1に示した音声評価部103の構成について説明する。図2は、本実施例1に係る音声評価部の構成を示す機能ブロック図である。図2に示すように、この音声評価部103は、ピッチ検出部110、分布算出部111、広がり算出部112、記憶部113、推定部114、評価部115を有する。

[0028] ピッチ検出部110は、音声ファイルの音声信号を周波数解析することで、ピッチ周波数を検出する処理部である。ピッチ周波数は、音声信号の基本周波数に対応するものである。ピッチ検出部110は、ピッチ周波数の情報を、分布算出部111に出力する。

[0029] ピッチ検出部110は、音声ファイルの音声信号から、発話区間を検出し、発話区間の音声信号に基づいて、ピッチ周波数を検出しても良い。たとえば、ピッチ検出部110は、音声信号の強さが閾値以上となる時刻を発話区間の開始時刻とする。ピッチ検出部110は、開始時刻以降において、音声信号の強さが閾値未満となる時刻を発話区間の終了時刻とする。ピッチ検出部110は、開始時刻から終了時刻までの区間を、発話区間とする。

[0030] ピッチ検出部110は、発話区間の音声信号を、所定の時間幅となる複数の区間に分割し、区間毎の音声信号を周波数解析することで、区間毎のピッチ周波数を検出する。ピッチ検出部110は、発話区間に含まれる複数の区間毎のピッチ周波数を、分布算出部111に出力する。

[0031] たとえば、ピッチ検出部110は、文献(D.Talkin, "A Robust Algorithm for Pitch Tracking (RAPT)," in Speech Coding & Synthesis, W.B.Kleijn and K.K.Pailwal (Eds.), Elsevier, pp.495-518, 1995)に基づいて、ピッチ周波数を検出しても良い。

[0032] 分布算出部111は、所定区間内の複数のピッチ周波数を基にして、ヒス

トグラムを算出する処理部である。図3は、ヒストグラムの一例を示す図である。図3に示すヒストグラム20の横軸は、ピッチ周波数に対応する軸であり、縦軸は検出頻度に対応する軸である。このヒストグラム20は、ピッチ周波数の検出頻度の分布に対応するものである。

[0033] 分布算出部111は、下記の処理を実行することで、ヒストグラム20の中心部と裾部とを特定する。分布算出部111は、所定区間内の各ピッチ周波数の平均 μ を算出する。分布算出部111は、所定区間内の各ピッチ周波数の標準偏差 σ を算出する。たとえば、分布算出部111は、ヒストグラム20の中心部を「 $\mu - \sigma \sim \mu + \sigma$ 」とする。分布算出部111は、ヒストグラム20の裾部を「 $(\mu - \sigma) / 2 \sim (\mu + \sigma) / 2$ 」、「 $2 \times (\mu - \sigma) / 2 \sim 2 \times (\mu + \sigma)$ 」とする。

[0034] たとえば、図3において、ピッチ周波数の範囲Aが中心部となり、範囲B1、B2が裾部となる。なお、分布算出部111は、上記以外の処理により、ヒストグラム20の中心部および裾部を算出してもよい。分布算出部111は、ヒストグラム20の概形から中心の山の始点と終点との間を中心部として特定し、この中心部以外の範囲を裾部として特定してもよい。

[0035] 分布算出部111は、所定区間内の複数のピッチ周波数の情報あるいはヒストグラム20の情報と、中心部の情報と、裾部の情報とを、広がり算出部112に出力する。

[0036] 広がり算出部112は、ヒストグラム20を補正した上で、ヒストグラム20の広がりを算出する処理部である。下記に説明するように、補正したヒストグラム20の広がり、補正したピッチ周波数に基づく標準偏差に対応する。

[0037] たとえば、広がり算出部112は、所定区間内の複数のピッチ周波数を、中心部に対応するピッチ周波数と、裾部に対応するピッチ周波数に分類する。広がり算出部112は、中央部に対応する各ピッチ周波数に重み「 α 」をそれぞれ乗算することで、中心部の各ピッチ周波数を補正する。 α の値をたとえば「1」とするが、管理者が適宜変

更してもよい。

- [0038] 広がり算出部 112 は、裾部に対応する各ピッチ周波数に重み「 β 」をそれぞれ乗算することで、裾部の各ピッチ周波数を補正する。 β の値をたとえば「0.1」とするが、管理者が適宜変更してもよい。
- [0039] 広がり算出部 112 は、補正した中央部および裾部の各ピッチ周波数を基にして、ピッチ周波数の標準偏差を算出する。このように、補正した中央部および裾部の各ピッチ周波数を基にして算出されるピッチ周波数の標準偏差が、補正したヒストグラム 20 の広がりに対応するものとなる。広がり算出部 112 は、補正したヒストグラム 20 の広がりに対応する標準偏差の情報を、推定部 114 および評価部 115 に出力する。
- [0040] 記憶部 113 は、平常時分布テーブル 113a を有する。記憶部 113 は、RAM、ROM、フラッシュメモリなどの半導体メモリ素子や、HDDなどの記憶装置に対応する。
- [0041] 平常時分布テーブル 113a は、平常時における話者のヒストグラムの広がり的情報を保持するテーブルである。たとえば、平常時分布テーブル 113a は、話者情報と、標準偏差とを対応づける。平常時分布テーブル 113a の標準偏差は、後述する推定部 114 によって推定される。
- [0042] 推定部 114 は、話者の平常時における標準偏差を推定する処理部である。推定部 114 は、話者情報と、平常時における標準偏差とを対応づけて、平常時分布テーブル 113a に格納する。推定部 114 は、音声ファイルに添付されている話者情報を取得する。
- [0043] 推定部 114 は、話者情報を取得し、話者情報に対応する標準偏差が平常時分布テーブル 113a に格納されていない場合には、「初期動作」を行う。一方、推定部 114 は、話者情報に対応する標準偏差が、平常時分布テーブル 113a に格納されている場合には、「更新動作」を行う。以下において、初期動作、更新動作について順に説明する。
- [0044] 推定部 114 が実行する初期動作について説明する。推定部 114 は、発話区間の開始時刻から、所定時間後（1分後）までの区間を初期区間とし、

初期区間における標準偏差を、広がり算出部 112 から取得する。初期区間における標準偏差は、上記のように、重み α 、 β により、ピッチ周波数を補正した後に、算出される標準偏差である。

[0045] なお、推定部 114 が、初期区間における標準偏差を算出しても良い。すなわち、推定部 114 は、初期区間内の複数のピッチ周波数を、中心部に対応するピッチ周波数と、裾部に対応するピッチ周波数に分類する。推定部 114 は、中央部に対応する各ピッチ周波数に重み「 α 」をそれぞれ乗算することで、中心部の各ピッチ周波数を補正する。推定部 114 は、裾部に対応する各ピッチ周波数に重み「 β 」をそれぞれ乗算することで、裾部の各ピッチ周波数を補正する。推定部 114 は、補正した中央部および裾部の各ピッチ周波数を基にして、ピッチ周波数の標準偏差を算出する。

[0046] 推定部 114 は、上記のように初期動作を行い、話者情報と、初期区間の標準偏差とを対応づけて、平常時分布テーブル 113 a に登録する。

[0047] 推定部 114 が実行する更新動作について説明する。推定部 114 は、平常時分布テーブル 113 a から、話者情報に対応する標準偏差を取得する。また、推定部 114 は、広がり算出部 112 から、所定区間内の標準偏差を取得する。以下の説明では、平常時分布テーブル 113 a から取得した、標準偏差を「平常時標準偏差」と表記し、広がり算出部 112 から取得した標準偏差を「評価標準偏差」と表記する。

[0048] 推定部 114 は、式 (1) に基づいて、新たな平常時標準偏差を算出し、算出した平常時標準偏差により、平常時分布テーブル 113 a の平常時標準偏差を更新する。

[0049] 平常時標準偏差 = $0.8 \times$ 平常時標準偏差 + $0.2 \times$ 評価標準偏差 \dots
(1)

[0050] 推定部 114 は、広がり算出部 112 から、所定区間内の標準偏差（評価標準偏差）を受信する度に、上記処理を繰り返し実行し、平常時分布テーブル 113 a を更新する。

[0051] なお、推定部 114 は、評価部 115 から、話者情報を取得し、平常時標

標準偏差の要求を受け付けた場合に、話者情報に対応する平常時標準偏差を平常時分布テーブル 113 a から取得し、平常時標準偏差を、評価部 115 に出力する。

[0052] 評価部 115 は、平常時標準偏差と、評価標準偏差とを基にして、話者の音声の印象を評価する処理部である。たとえば、評価部 115 は、音声ファイルに添付された話者情報を、推定部 114 に出力して、平常時標準偏差を取得する。評価部 115 は、評価標準偏差を、広がり算出部 112 から取得する。

[0053] 評価部 115 は、評価標準偏差が平常時標準偏差よりも大きい場合には、印象が良いと評価する。評価部 115 は、評価標準偏差が平常時標準偏差と同等である場合には、印象が普通であると判定する。評価部 115 は、評価標準偏差が平常時標準偏差より小さい場合には、印象がよいと判定する。

[0054] また、評価部 115 は、下記の様に評価を行ってもよい。たとえば、評価部 115 は、評価標準偏差が平常時標準偏差よりも大きく、評価標準偏差が平常時標準偏差との差が閾値以上である場合に、印象が良いと評価する。評価部 115 は、評価標準偏差が平常時標準偏差との差が閾値未満である場合に、印象が普通であると評価する。評価部 115 は、評価標準偏差が平常時標準偏差よりも小さく、評価標準偏差が平常時標準偏差との差が閾値以上である場合に、印象が悪いと評価する。

[0055] 評価部 115 は、話者情報と、評価結果の情報とを、評価結果格納部 104 に出力する。

[0056] 次に、本実施例 1 に係る音声評価装置 100 の音声評価部 103 の処理手順について説明する。図 4 は、本実施例 1 に係る音声評価部の処理手順を示すフローチャートである。図 4 に示すように、この音声評価部 103 のピッチ検出部 110 は、音声信号を受信する（ステップ S101）。

[0057] ピッチ検出部 110 は、音声信号を解析して、ピッチ周波数を算出する（ステップ S102）。音声評価部 103 の分布算出部 111 は、ピッチ周波数の分布を算出する（ステップ S103）。ステップ S103 において、分

布算出部 111 は、ピッチ周波数の分布を算出する処理は、上記のように、各ピッチ周波数に基づいて、ヒストグラム 20 の中心部および裾部を算出する処理に対応する。

[0058] 音声評価部 103 の広がり算出部 112 は、分布の広がりを算出する（ステップ S104）。ステップ S104 において、分布の広がりを算出する処理は、上記の評価標準偏差を算出する処理に対応する。

[0059] 音声評価部 103 の推定部 114 は、平常時の分布の広がりを算出する（ステップ S105）。平常時の分布の広がりを算出する処理は、上記の平常時標準偏差を算出する処理に対応する。

[0060] 音声評価部 103 の評価部 115 は、平常時標準偏差と評価標準偏差とを基にして、音声を評価する（ステップ S106）。

[0061] 音声評価部 103 は、音声終了の場合には（ステップ S107, Yes）、処理を終了する。一方、音声評価部 103 は、音声終了でない場合には（ステップ S107, No）、分析開始位置を更新し（ステップ S108）、ステップ S102 に移行する。

[0062] 次に、本実施例 1 に係る音声評価装置 100 の効果について説明する。音声評価装置 100 は、音声信号を解析して、ピッチ周波数の検出頻度の分布を作成し、分布の裾部分を抑える補正を行い、補正後の分布に基づき、音声の評価を行う。このため、ピッチ周波数を検出する処理により、仮に、半ピッチや倍ピッチを算出した場合にでも、これらのピッチは、裾部のピッチ周波数に分類され、影響を抑えた後に、音声の評価を行うことができ、音声を正確に評価することができる。たとえば、実際に明るい声であるにもかかわらず、半ピッチや倍ピッチが誤って算出された場合でも、暗い声と評価することを抑止することができる。

[0063] 音声評価装置 100 は、分布（ヒストグラム）の中心部に対応するピッチ周波数に重み α を乗算し、分布の裾部分に対応するピッチ周波数に重み β を乗算することで、分布を補正する処理を実行するため、半ピッチや倍ピッチ

の影響による誤り判定を抑止することができる。

[0064] 音声評価装置100の推定部114は、話者の平常時の分布の広がり推定し、評価部115が、平常時の分布の広がり、現在の分布の広がりと比較して、音声の印象を評価する。このため、話者自身の平常時の分布の広がりに基づいた音声の評価を行うことができる。

[0065] 音声評価装置114の推定部は、話者情報に対応する平常時の分布の広がり（標準偏差）が、平常時分布テーブル113aに格納されている場合には、「更新動作」を行う。これにより、話者に関する平常時の分布をより適切なものに補正することができる。

[0066] なお、音声評価装置100は、一例として、分布（ヒストグラム）の中心部に対応するピッチ周波数に重み α を乗算し、分布の裾部分に対応するピッチ周波数に重み β を乗算することで、分布を補正する場合について説明した。この処理は、分布に含まれる裾部のピッチ周波数の影響を取り除くという点において、分布に含まれる中央部のピッチ周波数を選択し、選択したピッチ周波数の標準偏差を基にして、音声の評価しているとも言える。

[0067] 音声評価装置100のピッチ検出部110は、発話区間の音声信号を、所定の時間幅となる複数の区間に分割し、区間毎の音声信号を周波数解析することで、区間毎のピッチ周波数を検出していたが、これに限定されるものではない。たとえば、ピッチ検出部110は、所定数の区間から検出した各ピッチ周波数の平均値を、ピッチ周波数として分布算出部111に出力してもよい。ピッチ検出部110は、所定数の区間から検出した各ピッチ周波数の上限値を、ピッチ周波数として分布算出部111に出力してもよい。ピッチ検出部110は、所定数の区間から検出した各ピッチ周波数の下限値を、ピッチ周波数として分布算出部111に出力してもよい。これにより、話者や環境に合わせて、ピッチ周波数を利用することができる。

[0068] 音声評価装置100の広がり算出部112は、発話区間に含まれる各ピッチ周波数の標準偏差を分布の広がりとして算出していたが、これに限定されるものではない。

い。たとえば、広がり算出部 112 は、発話区間に含まれる各ピッチ周波数の分散、範囲、平方和、四分位範囲のいずれかを、分布の広がりとして算出してもよい。

[0069] ところで、本実施例 1 に係る音声評価装置 100 は、以下に説明するように、発話区間のヒストグラムの概形を基にして、話者の音声を評価してもよい。図 5 は、音声評価装置のその他の処理を説明するための図である。図 5 に示すように、音声評価装置 100 は、発話区間の複数のピッチ周波数を基にして、ヒストグラム 25 を算出する。ヒストグラム 25 の横軸は、ピッチ周波数に対応する軸であり、縦軸は検出頻度に対応する軸である。

[0070] 音声評価装置 100 は、パターンマッチングなどを行って、ヒストグラム 25 の概形における中心部 C と、裾部 D1, D2 とを特定する。音声評価装置 100 は、中心部 C の検出頻度に重み α を乗算し、裾部 D1, D2 の検出頻度に重み β を乗算することで、ヒストグラム 25 を補正する。補正後のヒストグラムをヒストグラム 26 と表記する。たとえば、重み α を「1」とし、重み β を「0.1」とする。

[0071] 音声評価装置 100 は、補正後のヒストグラム 26 の広がり的大小を基にして、話者の音声の印象を判定する。たとえば、音声評価装置 100 は、ヒストグラム 26 の検出頻度が所定頻度以上となる範囲を、ヒストグラム 26 の広がりとして特定する。音声評価装置 100 は、ヒストグラム 26 の広がり、所定の閾値以上である場合には、話者の印象が明るいと評価する。一方、音声評価装置 100 は、ヒストグラム 26 の広がり、所定の閾値以上である場合には、話者の印象が暗いと評価する。

[0072] たとえば、半ピッチ／倍ピッチは分布（ヒストグラム 25）の裾部に存在しているため、裾部の重みを小さくし、中心部の重みを大きくして、ヒストグラム 25 を補正することで、誤って評価することを抑止することができる。

実施例 2

[0073] 図 6 は、本実施例 2 に係る音声評価システムの構成を示す図である。図 6

に示すように、この音声評価システムは、携帯端末 2 a、端末装置 2 b、分岐コネクタ 3、収録機器 150、クラウド 160 を有する。携帯端末 2 a は、電話網 15 a を介して、分岐コネクタ 3 に接続される。端末装置 2 b は、分岐コネクタ 3 に接続される。分岐コネクタ 3 は、収録機器 150 に接続される。収録機器 150 は、インターネット網 15 b を介して、クラウド 160 に接続される。たとえば、クラウド 160 には、音声評価装置 200 が含まれる。図示を省略するが、音声評価装置 200 は、複数のサーバによって構成されていてもよい。携帯端末 2 a および端末装置 2 b は、マイク（図示略）に接続される。

[0074] 話者 1 a による音声は、携帯端末 2 a のマイクにより集音され、集音された音声信号は、分岐コネクタ 3 を介して、収録機器 150 に送信される。以下の説明では、話者 1 a の音声信号を、「第 1 音声信号」と表記する。

[0075] 携帯端末 2 a は、第 1 音声信号に、話者 1 a の属性情報を付与する。たとえば、属性情報は、性別情報と、声の高さ情報とを含む。性別は、話者の性別を一意に識別するものである。声の高さか情報は、話者の声が高いか低いかを示す情報である。たとえば、話者 1 a は、自身の属性情報を、携帯端末 2 a に登録しておく。

[0076] 話者 1 b による音声は、端末装置 2 b のマイクにより集音され、集音された音声信号は、分岐コネクタ 3 を介して、収録機器 150 に送信される。以下の説明では、話者 1 b の音声信号を、「第 2 音声信号」と表記する。

[0077] 端末装置 2 b は、第 2 音声信号に、話者 1 b の属性情報を付与する。たとえば、話者 1 b は、自身の属性情報を、端末装置 2 b に登録しておく。話者 1 b の属性情報に関する説明は、話者 1 a の属性情報に関する説明と同様である。

[0078] 収録機器 150 は、第 1 音声信号および第 2 音声信号を収録する装置である。たとえば、収録機器 150 は、第 1 音声信号を受信すると、第 1 音声信号を、所定の音声ファイルフォーマットにより、音声ファイルに変換し、第 1 音声信号の音声ファイルを、音声評価装置 200 に送信する。第 1 音声信

号の音声ファイルには、話者1 aの属性情報が含まれる。以下の説明では、適宜、第1音声信号の音声ファイルを「第1音声ファイル」と表記する。

[0079] 収録機器150は、第2音声信号を受信すると、第2音声信号を、所定の音声ファイルフォーマットにより、音声ファイルに変換し、第2音声信号の音声ファイルを、音声評価装置200に送信する。第2音声信号の音声ファイルには、話者1 bの属性情報が含まれる。以下の説明では、適宜、第2音声信号の音声ファイルを「第2音声ファイル」と表記する。

[0080] 音声評価装置200は、第1音声ファイルの第1音声信号を基にして、話者1 aの音声の印象を評価する。音声評価装置200は、第2音声ファイルの第2音声信号を基にして、話者1 bの音声の印象を評価する。そして、音声評価装置200は、話者1 aの音声の印象の評価結果および話者2 aの音声の印象の評価結果を基にして、話者1 a、1 b間の会話全体のスコアを算出する。

[0081] 図7は、本実施例2に係る音声評価装置の構成を示す機能ブロック図である。図7に示すように、この音声評価装置200は、受信部201、記憶装置202、音声評価部203、評価結果格納部204を有する。

[0082] 受信部201は、収録機器150から、第1音声ファイルおよび第2音声ファイルを受信する処理部である。受信部201は、受信した第1音声ファイルおよび第2音声ファイルを、記憶部202の音声ファイルテーブル202 aに登録する。受信部201は、通信装置に対応する。

[0083] 記憶装置202は、音声ファイルテーブル202 aと、評価結果テーブル202 bを有する。記憶装置202は、RAM、ROM、フラッシュメモリなどの半導体メモリ素子や、HDDなどの記憶装置に対応する。

[0084] 音声ファイルテーブル202 aは、第1音声ファイルおよび第2音声ファイルを格納するテーブルである。

[0085] 評価結果テーブル202 bは、評価結果を格納するテーブルである。たとえば、評価結果テーブル202 bは、話者1 aの評価結果、話者1 bの評価結果、話者1 a、1 b間の会話全体のスコアを格納する。

- [0086] 音声評価部203は、第1音声ファイルおよび第2音声ファイルを基にして、話者1a、1bの音声の印象を評価する。そして、音声評価部203は、話者1a、1の音声の印象の評価結果を基にして、話者1a、1b間の会話全体のスコアを算出する。音声評価部203は、話者1a、1bの音声の印象の評価結果および会話全体のスコアを、評価結果格納部204に出力する。
- [0087] 評価結果格納部204は、話者1a、1の音声の印象の評価結果および会話全体のスコアを、評価結果テーブル202bに格納する処理部である。
- [0088] ここで、図7に示した音声評価部203、評価結果格納部204は、所定の制御部（図示略）に対応する。たとえば、制御部は、CPUやMPUなどによって実現できる。また、制御部は、ASICやFPGAなどのハードワイヤードロジックによっても実現できる。
- [0089] 続いて、図7に示した音声評価部203の構成について説明する。図8は、本実施例2に係る音声評価部の構成を示す機能ブロック図である。図8に示すように、この音声評価部203は、ピッチ検出部210、分布算出部211、広がり算出部212、記憶部213、推定部114、評価部115を有する。
- [0090] ピッチ検出部210は、音声ファイルの音声信号を周波数解析することで、所定区間毎のピッチ周波数を検出する処理部である。たとえば、ピッチ検出部210は、第1音声ファイルの第1音声信号を周波数解析することで、第1音声信号の第1ピッチ周波数を検出する。また、ピッチ検出部210は、第2音声ファイルの第2音声信号を周波数解析することで、第2音声信号の第2ピッチ周波数を検出する。ピッチ検出部210が、音声信号からピッチ周波数を検出する処理は、図2に示したピッチ検出部110が、音声信号からピッチ周波数を検出する処理と同様である。
- [0091] ピッチ検出部210は、話者1aの属性情報と、複数の第1ピッチ周波数とを分布算出部211に出力する。また、ピッチ検出部210は、話者1b

の属性情報と、複数の第2ピッチ周波数とを分布算出部211に出力する。

[0092] 分布算出部211は、所定区間内の複数のピッチ周波数を基にして、ヒストグラムを算出する処理部である。たとえば、分布算出部211は、所定区間内の複数の第1ピッチ周波数を基にして、第1ヒストグラムを算出する。分布算出部211は、所定区間内の複数の第2ピッチ周波数を基にして、第2ヒストグラムを算出する。分布算出部211が、ヒストグラムを算出する処理は、図2に示した分布算出部111が、ヒストグラムを算出する処理と同様である。

[0093] 分布算出部211は、複数の第1ピッチ周波数を基にして算出した、第1ヒストグラムの情報を、広がり算出部212に出力する。第1ヒストグラムの情報は、第1ヒストグラムの中心部の情報、第1ヒストグラムの裾部の情報を含む。

[0094] 分布算出部211は、複数の第2ピッチ周波数を基にして算出した、第2ヒストグラムの情報を、広がり算出部212に出力する。第2ヒストグラムの情報は、第2ヒストグラムの中心部の情報、第2ヒストグラムの裾部の情報を含む。

[0095] 広がり算出部212は、ヒストグラムを補正した上で、ヒストグラムの広がりを算出する処理部である。たとえば、広がり算出部212は、第1ヒストグラムを補正した上で、第1ヒストグラムの広がりを算出する。広がり算出部212は、第2ヒストグラムを補正した上で、第2ヒストグラムの広がりを算出する。広がり算出部212が、ヒストグラムの広がりを算出する処理は、図2に示した広がり算出部112が、ヒストグラムの広がりを算出する処理と同様である。たとえば、補正したヒストグラムの広がりには、補正したピッチ周波数（第1ピッチ周波数、第2ピッチ周波数）に基づく標準偏差に対応する。

[0096] 広がり算出部212は、属性情報と対応づけて、補正したヒストグラムの広がりに対応する標準偏差の情報を、推定部214および評価部215に出力する。たとえば、広がり算出部212は、話者1aの属性情報と対応づけ

て、補正した第1ヒストグラムの広がりに対応する標準偏差の情報を、推定部214および評価部215に出力する。広がり算出部212は、話者1bの属性情報と対応づけて、補正した第2ヒストグラムの広がりに対応する標準偏差の情報を、推定部214および評価部215に出力する。

[0097] 記憶部213は、平常時分布テーブル213aを有する。記憶部213は、RAM、ROM、フラッシュメモリなどの半導体メモリ素子や、HDDなどの記憶装置に対応する。

[0098] 平常時分布テーブル213aは、属性情報に対応する話者の平常時における、ヒストグラムの広がり情報を保持するテーブルである。たとえば、平常時分布テーブル213aは、属性情報と、標準偏差とを対応づける。平常時分布テーブル213aの標準偏差は、後述する推定部214によって推定される。

[0099] 図9は、本実施例2に係る平常時分布テーブルのデータ構造の一例を示す図である。図9に示すように、この平常時分布テーブル213aは、属性情報と、標準偏差とを対応づける。属性情報は、性別情報と、声の高さ情報とを対応づける。

[0100] 推定部214は、属性情報毎に、話者の平常時における標準偏差を推定する処理部である。推定部214は、属性情報と、平常時における標準偏差とを対応づけて、平常時分布テーブル213aに格納する。推定部214は、音声ファイル（第1音声ファイルおよび第2音声ファイル）に添付されている属性情報を取得する。

[0101] 推定部214は、属性情報を取得し、属性情報に対応する標準偏差が平常時分布テーブル213aに格納されていない場合には、「初期動作」を行う。一方、推定部214は、属性情報に対応する標準偏差が、平常時分布テーブル213aに格納されている場合には、「更新動作」を行う。推定部214による、初期動作、更新動作に関する説明は、属性情報毎に、実行される点を除いて、推定部114の初期動作、更新動作に関する説明と同様である。

- 。
- [0102] 以下の説明では、話者 1 a の属性情報に対応する、平常時分布テーブル 2 1 3 a から取得した、標準偏差を「第 1 平常時標準偏差」と表記する。話者 1 a の属性情報に対応する、広がり算出部 2 1 2 から取得した標準偏差を「第 1 評価標準偏差」と表記する。
- [0103] 話者 1 b の属性情報に対応する、平常時分布テーブル 2 1 3 a から取得した、標準偏差を「第 2 平常時標準偏差」と表記する。話者 1 b の属性情報に対応する、広がり算出部 2 1 2 から取得した標準偏差を「第 2 評価標準偏差」と表記する。
- [0104] なお、推定部 2 1 4 は、評価部 2 1 5 から、話者 1 a の属性情報を取得し、第 1 平常時標準偏差の要求を受け付けた場合に、第 1 平常時標準偏差を平常時分布テーブル 2 1 3 a から取得し、第 1 平常時標準偏差を、評価部 2 1 5 に出力する。
- [0105] 推定部 2 1 4 は、評価部 2 1 5 から、話者 1 b の属性情報を取得し、第 2 平常時標準偏差の要求を受け付けた場合に、第 2 平常時標準偏差を平常時分布テーブル 2 1 3 a から取得し、第 2 平常時標準偏差を、評価部 2 1 5 に出力する。
- [0106] 評価部 2 1 5 は、各標準偏差を基にして、話者 1 a, 1 b の音声の印象を評価する処理部である。また、評価部 2 1 5 は、話者 1 a, 1 b の音声の印象の評価結果を基にして、話者 1 a、1 b 間の会話全体のスコアを算出する。
- 。
- [0107] 具体的に、評価部 2 1 5 は、第 1 平常時標準偏差と、第 1 評価標準偏差とを基にして、話者 1 a の音声の印象を評価する。また、評価部 2 1 5 は、第 2 平常時標準偏差と、第 2 評価標準偏差とを基にして、話者 1 b の音声の印象を評価する。評価部 2 1 5 が、各標準偏差を基にして、話者 1 a, 1 b の音声の印象を評価する処理は、評価部 1 1 5 が、音声の評価する処理と同様である。
- [0108] 評価部 2 1 5 は、上記処理を繰り返し実行することで、所定区間毎に、話

者 1 a の評価結果および話者 1 b の評価結果を特定する。

[0109] 続いて、評価部 215 は、話者 1 a の評価結果および話者 1 b の評価結果を特定した後に、式 (2) に基づいて、会話全体のスコアを算出する。

[0110] 会話全体のスコア = (話者 1 a の印象が良いと評価された区間数 + 話者 1 b の印象が良いと評価された区間数) / (全体の区間数 × 2) × 100 . . . (2)

[0111] 評価部 215 は、話者 1 a、1 b の音声の印象の評価結果および会話全体のスコアを、評価結果格納部 204 に出力する。

[0112] 次に、本実施例 2 に係る音声評価装置 200 の効果について説明する。音声評価装置 200 は、話者の属性情報に合わせて、平常時標準偏差を学習しておき、評価時において、話者の属性情報に応じた平常時標準偏差を用いて、話者の音声を評価する。このため、話者の属性情報に基づく、音声の評価を行うことができ、評価の正確性を向上させることができる。

[0113] また、音声評価装置 200 は、各話者 1 a、1 b の音声の評価結果に基づいて、会話全体のスコアを算出するため、管理者などが係るスコアを参照することで、会話全体の善し悪しを把握することができる。たとえば、式 (2) により算出される会話全体のスコアは、値が大きいほど、会話内容が良かったと言える。

実施例 3

[0114] 図 10 は、本実施例 3 に係る音声評価システムの構成を示す図である。図 10 に示すように、この音声評価システムは、マイク 30 A、30 B、30 C、収録機器 300、クラウド 170 を有する。マイク 30 A ~ 30 C は、収録機器 300 に接続される。収録機器 300 は、インターネット網 15 b を介して、クラウド 170 に接続される。たとえば、クラウド 170 には、音声評価装置 400 が含まれる。図示を省略するが、音声評価装置 400 は、複数のサーバによって構成されていてもよい。

[0115] 話者 1 A による音声は、マイク 30 a により集音され、集音された音声信号は、収録機器 300 に出力される。話者 1 B による音声は、マイク 30 b

により集音され、集音された音声信号は、収録機器300に出力される。話者1Cによる音声は、マイク30cにより集音され、集音された音声信号は、収録機器300に出力される。

[0116] 以下の説明では、話者1Aの音声信号を、「第1音声信号」と表記する。話者1Bの音声信号を、「第2音声信号」と表記する。話者1Cの音声信号を、「第3音声信号」と表記する。

[0117] なお、第1音声信号には、話者1Aの話者情報が付与される。話者情報は、話者を一意に識別する情報である。第2音声信号には、話者1Bの話者情報が付与される。第3音声信号には、話者1Cの話者情報が付与される。

[0118] 収録機器300は、第1音声信号、第2音声信号、第3音声信号を収録する装置である。また、収録装置300は、各音声信号のピッチ周波数を検出する処理を実行する。収録装置300は、話者情報と、所定区間毎のピッチ周波数とを対応づけて、音声評価装置400に送信する。

[0119] 音声評価装置400は、収録機器300から受信する各話者情報のピッチ周波数を基にして、各話者の音声を評価する処理部である。また、音声評価装置400は、各話者の音声の評価結果を基にして、話者1A~1Cの会話の印象を評価する。

[0120] 図11は、本実施例3に係る収録機器の構成を示す機能ブロック図である。図11に示すように、この収録機器300は、AD変換部310a~310bと、ピッチ検出部320と、ファイル化部330と、送信部340とを有する。

[0121] AD変換部310aは、マイク30aから第1音声信号を受信し、AD変換を実行する処理部である。具体的には、AD変換部310aは、第1音声信号（アナログ信号）を、第1音声信号（デジタル信号）に変換する。AD変換部310aは、第1音声信号（デジタル信号）を、ピッチ検出部320に出力する。以下の説明では、AD変換部310aから出力される第1音声信号（デジタル信号）を単に第1音声信号と表記する。

- [0122] AD変換部310bは、マイク30bから第2音声信号を受信し、AD変換を実行する処理部である。具体的には、AD変換部310bは、第2音声信号（アナログ信号）を、第2音声信号（デジタル信号）に変換する。AD変換部310bは、第2音声信号（デジタル信号）を、ピッチ検出部320に出力する。以下の説明では、AD変換部310bから出力される第2音声信号（デジタル信号）を単に第2音声信号と表記する。
- [0123] AD変換部310cは、マイク30cから第3音声信号を受信し、AD変換を実行する処理部である。具体的には、AD変換部310cは、第3音声信号（アナログ信号）を、第3音声信号（デジタル信号）に変換する。AD変換部310cは、第3音声信号（デジタル信号）を、ピッチ検出部320に出力する。以下の説明では、AD変換部310cから出力される第3音声信号（デジタル信号）を単に第3音声信号と表記する。
- [0124] ピッチ検出部320は、音声信号を周波数解析することで、所定区間毎のピッチ周波数を算出する処理部である。たとえば、ピッチ検出部320は、第1音声信号を周波数解析することで、第1音声信号の第1ピッチ周波数を検出する。ピッチ検出部320は、第2音声信号を周波数解析することで、第2音声信号の第2ピッチ周波数を検出する。ピッチ検出部320は、第3音声信号を周波数解析することで、第3音声信号の第3ピッチ周波数を検出する。
- [0125] ピッチ検出部320は、話者1Aの話者情報と、所定区間毎の第1ピッチ周波数とを対応づけて、ファイル化部330に出力する。ピッチ検出部320は、話者1Bの話者情報と、所定区間毎の第2ピッチ周波数とを対応づけて、ファイル化部330に出力する。ピッチ検出部320は、話者1Cの話者情報と、所定区間毎の第3ピッチ周波数とを対応づけて、ファイル化部330に出力する。
- [0126] ファイル化部330は、ピッチ検出部320から受け付ける情報をファイル化することで、「音声ファイル情報」を生成する処理部である。この音声ファイル情報には、話者情報と、所定区間毎のピッチ周波数とを対応づけた

情報を含む。具体的に、音声ファイル情報は、話者1Aの話者情報と、所定区間毎の第1ピッチ周波数とを対応づけた情報を含む。音声ファイル情報は、話者1Bの話者情報と、所定区間毎の第2ピッチ周波数とを対応づけた情報を含む。音声ファイル情報は、話者1Cの話者情報と、所定区間毎の第3ピッチ周波数とを対応づけた情報を含む。ファイル化部330は、音声ファイル情報を、送信部340に出力する。

[0127] 送信部340は、ファイル化部330から音声ファイル情報を取得し、取得した音声ファイル情報を、音声評価装置400に送信する。

[0128] 図12は、本実施例3に係る音声評価装置の構成を示す機能ブロック図である。図12に示すように、この音声評価装置400は、受信部401、記憶装置402、音声評価部403、評価結果格納部404を有する。

[0129] 受信部401は、収録機器300から、音声ファイル情報を受信する処理部である。受信部401は、受信した音声ファイル情報を、記憶部402の音声ファイルテーブル402aに登録する。受信部401は、通信装置に対応する。

[0130] 記憶装置402は、音声ファイルテーブル402aと、評価結果テーブル402bを有する。記憶装置402は、RAM、ROM、フラッシュメモリなどの半導体メモリ素子や、HDDなどの記憶装置に対応する。

[0131] 音声ファイルテーブル402aは、音声ファイル情報を格納するテーブルである。音声ファイル情報は、話者1Aの話者情報と、所定区間毎の第1ピッチ周波数とを対応づけた情報を含む。音声ファイル情報は、話者1Bの話者情報と、所定区間毎の第2ピッチ周波数とを対応づけた情報を含む。音声ファイル情報は、話者1Cの話者情報と、所定区間毎の第3ピッチ周波数とを対応づけた情報を含む。

[0132] 評価結果テーブル402bは、評価結果を格納するテーブルである。たとえば、評価結果テーブル402bは、話者1A~1Cの評価結果、話者1A~1C間の会話全体の評価結果を格納する。

[0133] 音声評価部403は、音声ファイル情報を基にして、話者1A~1Cの音

声の印象を評価する。そして、音声評価部403は、話者1A～1Cの音声の印象の評価結果を基にして、話者1A～1C間の会話全体を評価する。音声評価部403は、話者1A～1Cの音声の印象の評価結果および会話全体の評価結果を、評価結果格納部404に出力する。

[0134] 評価結果格納部404は、話者1A～1Cの音声の印象の評価結果および会話全体の評価結果を、評価結果テーブル402bに格納する処理部である。

[0135] ここで、図12に示した音声評価部403、評価結果格納部404は、所定の制御部（図示略）に対応する。たとえば、制御部は、CPUやMPUなどによって実現できる。また、制御部は、ASICやFPGAなどのハードワイヤードロジックによっても実現できる。

[0136] 続いて、図13に示した音声評価部403の構成について説明する。図13は、本実施例3に係る音声評価部の構成を示す機能ブロック図である。図13に示すように、この音声評価部403は、ピッチ取得部410、分布算出部411、広がり算出部412、記憶部413、推定部414、評価部415を有する。

[0137] ピッチ取得部410は、音声ファイルテーブル402aから音声ファイル情報を取得する処理部である。ピッチ取得部410は、音声ファイル情報を、分布算出部411に出力する。

[0138] 分布算出部411は、所定区間内の複数のピッチ周波数を基にして、ヒストグラムを算出する処理部である。たとえば、分布算出部411は、所定区間内の複数の第1ピッチ周波数を基にして、第1ヒストグラムを算出する。分布算出部411は、所定区間内の複数の第2ピッチ周波数を基にして、第2ヒストグラムを算出する。分布算出部411は、所定区間内の複数の第3ピッチ周波数を基にして、第3ヒストグラムを算出する。分布算出部411が、ヒストグラムを算出する処理は、図2に示した分布算出部111が、ヒストグラムを算出する処理と同様である。

- [0139] 分布算出部411は、複数の第1ピッチ周波数を基にして算出した、第1ヒストグラムの情報を、広がり算出部412に出力する。第1ヒストグラムの情報は、第1ヒストグラムの中心部の情報、第1ヒストグラムの裾部の情報を含む。
- [0140] 分布算出部411は、複数の第2ピッチ周波数を基にして算出した、第2ヒストグラムの情報を、広がり算出部412に出力する。第2ヒストグラムの情報は、第2ヒストグラムの中心部の情報、第2ヒストグラムの裾部の情報を含む。
- [0141] 分布算出部411は、複数の第3ピッチ周波数を基にして算出した、第3ヒストグラムの情報を、広がり算出部412に出力する。第2ヒストグラムの情報は、第3ヒストグラムの中心部の情報、第3ヒストグラムの裾部の情報を含む。
- [0142] 広がり算出部412は、ヒストグラムを補正した上で、ヒストグラムの広がりを算出する処理部である。たとえば、広がり算出部412は、第1ヒストグラムを補正した上で、第1ヒストグラムの広がりを算出する。広がり算出部412は、第2ヒストグラムを補正した上で、第2ヒストグラムの広がりを算出する。広がり算出部412は、第3ヒストグラムを補正した上で、第3ヒストグラムの広がりを算出する。広がり算出部412が、ヒストグラムの広がりを算出する処理は、図2に示した広がり算出部112が、ヒストグラムの広がりを算出する処理と同様である。たとえば、補正したヒストグラムの広がりには、補正したピッチ周波数（第1ピッチ周波数、第2ピッチ周波数、第3ピッチ周波数）に基づく標準偏差に対応する。
- [0143] 広がり算出部412は、話者情報と対応づけて、補正したヒストグラムの広がりに対応する標準偏差の情報を、推定部414および評価部415に出力する。たとえば、広がり算出部412は、話者1Aの話者情報と対応づけて、補正した第1ヒストグラムの広がりに対応する標準偏差の情報を、推定部414および評価部415に出力する。広がり算出部412は、話者1Bの話者情報と対応づけて、補正した第2ヒストグラムの広がりに対応する標

標準偏差の情報を、推定部414および評価部415に出力する。広がり算出部412は、話者1Cの話者情報と対応づけて、補正した第2ヒストグラムの広がりに対応する標準偏差の情報を、推定部414および評価部415に出力する。

[0144] 記憶部413は、平常時分布テーブル413aを有する。記憶部413は、RAM、ROM、フラッシュメモリなどの半導体メモリ素子や、HDDなどの記憶装置に対応する。

[0145] 平常時分布テーブル413aは、話者情報に対応する話者の平常時における、ヒストグラムの広がり情報を保持するテーブルである。たとえば、平常時分布テーブル413aは、話者情報と、標準偏差とを対応づける。平常時分布テーブル413aの標準偏差は、後述する推定部414によって推定される。

[0146] 推定部414は、話者情報毎に、話者の平常時における標準偏差を推定する処理部である。推定部414は、話者情報と、平常時における標準偏差とを対応づけて、平常時分布テーブル413aに格納する。推定部414は、音声ファイル情報に添付されている話者情報を取得する。

[0147] 推定部414は、話者情報を取得し、話者情報に対応する標準偏差が平常時分布テーブル413aに格納されていない場合には、「初期動作」を行う。一方、推定部414は、話者情報に対応する標準偏差が、平常時分布テーブル413aに格納されている場合には、「更新動作」を行う。推定部414による、初期動作、更新動作に関する説明は、図2に示した推定部114の初期動作、更新動作に関する説明と同様である。

[0148] 以下の説明では、話者1Aの話者情報に対応する、平常時分布テーブル413aから取得した、標準偏差を「第1平常時標準偏差」と表記する。話者1Aの話者情報に対応する、広がり算出部412から取得した標準偏差を「第1評価標準偏差」と表記する。

[0149] 話者1Bの話者情報に対応する、平常時分布テーブル413aから取得した、標準偏差を「第2平常時標準偏差」と表記する。話者1Bの話者情報に

対応する、広がり算出部412から取得した標準偏差を「第2評価標準偏差」と表記する。

[0150] 話者1Cの話者情報に対応する、平常時分布テーブル413aから取得した、標準偏差を「第3平常時標準偏差」と表記する。話者1Cの話者情報に対応する、広がり算出部412から取得した標準偏差を「第3評価標準偏差」と表記する。

[0151] なお、推定部414は、評価部415から、話者1Aの話者情報を取得し、第1平常時標準偏差の要求を受け付けた場合に、第1平常時標準偏差を平常時分布テーブル413aから取得し、第1平常時標準偏差を、評価部415に出力する。

[0152] 推定部414は、評価部415から、話者1Bの話者情報を取得し、第2平常時標準偏差の要求を受け付けた場合に、第2平常時標準偏差を平常時分布テーブル413aから取得し、第2平常時標準偏差を、評価部415に出力する。

[0153] 推定部414は、評価部415から、話者1Cの話者情報を取得し、第3平常時標準偏差の要求を受け付けた場合に、第3平常時標準偏差を平常時分布テーブル413aから取得し、第3平常時標準偏差を、評価部415に出力する。

[0154] 評価部415は、各標準偏差を基にして、話者1A～1Cの音声の印象を評価する処理部である。また、評価部415は、話者1A～1Cの音声の印象の評価結果を基にして、話者1A～1Cの会話全体の評価を行う。

[0155] 具体的に、評価部415は、第1平常時標準偏差と、第1評価標準偏差とを基にして、話者1Aの音声の印象を評価する。評価部415は、第2平常時標準偏差と、第2評価標準偏差とを基にして、話者1Bの音声の印象を評価する。評価部415は、第3平常時標準偏差と、第3評価標準偏差とを基にして、話者1Cの音声の印象を評価する。評価部415が、各標準偏差を基にして、話者1A～1Cの音声の印象を評価する処理は、評価部115が、音声を評価する処理と同様である。

- [0156] 評価部415は、上記処理を繰り返し実行することで、所定区間毎に、話者1Aの評価結果、話者1Bの評価結果、話者1Cの評価結果を特定する。
- [0157] 続いて、評価部415は、話者1A～1Cの評価結果を特定した後に、会話全体の評価を行う。たとえば、評価部415は、話者毎に、所定フレーム間の各評価結果「良い、普通、悪い」のうち、平均の評価結果を特定する。たとえば、評価部415は、所定フレーム間の各評価結果のうち、もっとも多い評価結果を、平均の評価結果とする。
- [0158] 評価部415は、各話者1A～1Cの平均の評価結果が非常に近い場合には、良い会話であると判定する。評価部415は、各話者1A～1Cの平均の評価結果が異なる場合には、悪い会話であると判定する。
- [0159] たとえば、評価部415は、各話者1A～1Cの平均の評価結果を比較し、2以上の平均の評価結果が一致した場合には、良い会話であると判定する。一方、評価部415は、各話者1A～1Cの平均の評価結果を比較し、2以上の平均の評価結果が一致しない場合には、悪い会話であると判定する。
- [0160] 評価部415は、話者1A～1Cの音声の印象の評価結果および会話全体の評価結果を、評価結果格納部404に出力する。
- [0161] 次に、本実施例3に係る音声評価装置400の効果について説明する。音声評価装置400は、話者の話者情報に合わせて、平常時標準偏差を学習しておき、評価時において、話者の話者情報に応じた平常時標準偏差を用いて、話者の音声を評価する。このため、話者情報に基づく、音声の評価を行うことができ、評価の正確性を向上させることができる。
- [0162] また、音声評価装置400は、各話者1A～1Cの音声の評価結果に基づいて、会話全体を評価するため、管理者などが会話全体の善し悪しを把握することができる。
- [0163] また、本実施例3に係る音声評価装置400は、会話全体を評価する場合に、式(3)に基づいて、会話全体のスコアを算出しても良い。
- [0164] 会話全体のスコア = (話者1Aの印象が良いと評価された区間数 + 話者1

Bの印象が良いと評価された区間数+話者1 Cの印象が良いと評価された区間数) / (全体の区間数×3) × 100・・・(3)

[0165] 次に、上記実施例に示した音声評価装置100、200、400と同様の機能を実現するコンピュータのハードウェア構成の一例について説明する。図14は、音声評価装置と同様の機能を実現するコンピュータのハードウェア構成の一例を示す図である。

[0166] 図14に示すように、コンピュータ500は、各種演算処理を実行するCPU501と、ユーザからのデータの入力を受け付ける入力装置502と、ディスプレイ503とを有する。また、コンピュータ500は、記憶媒体からプログラム等を読み取る読み取り装置504と、有線または無線ネットワークを介して収録機器等との間でデータの授受を行うインターフェース装置505とを有する。また、コンピュータ500は、各種情報を一時記憶するRAM506と、ハードディスク装置507とを有する。そして、各装置501～507は、バス508に接続される。

[0167] ハードディスク装置507は、ピッチ検出プログラム507a、分布算出プログラム507b、広がり算出プログラム507c、推定プログラム507d、評価プログラム507eを有する。CPU501は、ピッチ検出プログラム507a、分布算出プログラム507b、広がり算出プログラム507c、推定プログラム507d、評価プログラム507dを読み出してRAM506に展開する。

[0168] ピッチ検出プログラム507aは、ピッチ検出プロセス506aとして機能する。分布算出プログラム507bは、分布算出プロセス506bとして機能する。広がり算出プログラム507cは、広がり算出プロセス506cとして機能する。推定プログラム507dは、推定プロセス506dとして機能する。評価プログラム507eは、評価プロセス506eとして機能する。

[0169] ピッチ検出プロセス506aの処理は、ピッチ検出部110、210、320の処理に対応する。分布算出プロセス506bは、分布算出部111、

2 1 1, 4 1 1 の処理に対応する。広がり算出プロセス 5 0 6 c の処理は、広がり算出部 1 1 2, 2 1 2, 4 1 2 の処理に対応する。推定プロセス 5 0 6 d は、推定部 1 1 4, 2 1 4, 4 1 4 の処理に対応する。

[0170] なお、各プログラム 5 0 7 a ~ 5 0 7 e については、必ずしも最初からハードディスク装置 5 0 7 に記憶させておかなくても良い。例えば、コンピュータ 5 0 0 に挿入されるフレキシブルディスク (FD)、CD-ROM、DVD ディスク、光磁気ディスク、IC カードなどの「可搬用の物理媒体」に各プログラムを記憶させておく。そして、コンピュータ 5 0 0 が各プログラム 5 0 7 a ~ 5 0 7 e を読み出して実行するようにしても良い。

[0171] 以上の各実施例を含む実施形態に関し、さらに以下の付記を開示する。

[0172] (付記 1) 音声信号を解析してピッチ周波数を検出し、

検出した前記ピッチ周波数の検出頻度の分布に基づいて、検出した前記ピッチ周波数のうち評価対象とする評価対象領域を選択し、

前記検出頻度の分布と選択した前記評価対象領域に基づいて、音声の評価する

処理をコンピュータに実行させることを特徴とする音声評価プログラム。

[0173] (付記 2) 前記分布の中心部に対応するピッチ周波数に第 1 の重みを乗算し、前記分布の裾部分に対応するピッチ周波数に前記第 1 の重みよりも小さい第 2 の重みを乗算することで、前記分布を補正する処理を更に実行させ、前記評価する処理は、補正した分布の広がりを基にして、前記発話区間内の音声の印象を評価することを特徴とする付記 1 に記載の音声評価プログラム。

[0174] (付記 3) 前記ピッチ周波数を検出する処理は、各時刻のピッチ周波数と、一定時間内における平均のピッチ周波数、上限のピッチ周波数、下限のピッチ周波数のいずれかを検出することを特徴とする付記 1 または 2 に記載の音声評価プログラム。

[0175] (付記 4) 前記評価する処理は、複数のピッチ周波数に関する、分散、標準偏差、範囲、平方和、四分位範囲のいずれかの統計量を算出し、前記統計量を基にして、音声の印象を評価することを特徴とする付記 1、2 または 3 に

記載の音声評価プログラム。

- [0176] (付記5) 予め定めた所定区間内の音声信号を解析して得られるピッチ周波数を基にして、ユーザの平常時の検出頻度の基準分布を推定する処理を更に実行させ、前記評価する処理は、前記基準分布と、補正された分布とを基にして、音声の印象を評価することを特徴とする付記2に記載の音声評価プログラム。
- [0177] (付記6) 前記基準分布と、話者情報とを対応づけて記憶装置に記憶する処理を更に実行させ、前記評価する処理は、話者情報に対応する基準分布を選択し、選択した基準分布の広がり、と、補正された分布の広がりとを基にして、音声の印象を評価することを特徴とする付記5に記載の音声評価プログラム。
- [0178] (付記7) 前記基準分布を推定する処理は、指定された話者情報に対応する基準分布が前記記憶装置に記憶されている場合に、記憶装置に記憶された指定された話者情報に対応する基準分布を、前記基準分布を推定する場合の初期値に設定することを特徴とする付記6に記載の音声評価プログラム。
- [0179] (付記8) 前記分布を補正する処理は、予め定められた所定区間における複数のピッチ周波数を基にして、前記分布の中心部および前記分布の裾部を特定することを特徴とする付記2に記載の音声評価プログラム。
- [0180] (付記9) 前記評価する処理は、異なる話者情報毎に音声の評価し、複数話者による会話全体を評価することを特徴とする付記1に記載の音声評価プログラム。
- [0181] (付記10) 前記評価する処理は、各時刻における音声の印象をそれぞれ評価し、各評価結果に関するスコアを算出することを特徴とする付記1に記載の音声評価プログラム。
- [0182] (付記11) 前記評価する処理は、前記分布の広がり大きい場合は声の印象が良いと評価し、前記分布の広がり小さい場合は声の印象が悪いと評価することを特徴とする付記2、3または4に記載の音声評価プログラム。
- [0183] (付記12) コンピュータが実行する音声評価方法であって、

音声信号を解析してピッチ周波数を検出し、

検出した前記ピッチ周波数の検出頻度の分布に基づいて、検出した前記ピッチ周波数のうち評価対象とする評価対象領域を選択し、

前記検出頻度の分布と選択した前記評価対象領域に基づいて、音声の評価する

処理をコンピュータに実行させることを特徴とする音声評価方法。

[0184] (付記 1 3) 前記分布の中心部に対応するピッチ周波数に第 1 の重みを乗算し、前記分布の裾部分に対応するピッチ周波数に前記第 1 の重みよりも小さい第 2 の重みを乗算することで、前記分布を補正する処理を更に実行させ、前記評価する処理は、補正した分布の広がりに基づいて、前記発話区間内の音声の印象を評価することを特徴とする付記 1 2 に記載の音声評価方法。

[0185] (付記 1 4) 前記ピッチ周波数を検出する処理は、各時刻のピッチ周波数と、一定時間内における平均のピッチ周波数、上限のピッチ周波数、下限のピッチ周波数のいずれかを検出することを特徴とする付記 1 2 または 1 3 に記載の音声評価方法。

[0186] (付記 1 5) 前記評価する処理は、複数のピッチ周波数に関する、分散、標準偏差、範囲、平方和、四分位範囲のいずれかの統計量を算出し、前記統計量を基にして、音声の印象を評価することを特徴とする付記 1 2、1 3 または 1 4 に記載の音声評価方法。

[0187] (付記 1 6) 予め定めた所定区間内の音声信号を解析して得られるピッチ周波数を基にして、ユーザの平常時の検出頻度の基準分布を推定する処理を更に実行させ、前記評価する処理は、前記基準分布と、補正された分布とを基にして、音声の印象を評価することを特徴とする付記 1 3 に記載の音声評価方法。

[0188] (付記 1 7) 前記基準分布と、話者情報とを対応づけて記憶装置に記憶する処理を更に実行させ、前記評価する処理は、話者情報に対応する基準分布を選択し、選択した基準分布の広がり、補正された分布の広がりとを基にして、音声の印象を評価することを特徴とする付記 1 6 に記載の音声評価方法

- 。
- [0189] (付記 18) 前記基準分布を推定する処理は、指定された話者情報に対応する基準分布が前記記憶装置に記憶されている場合に、記憶装置に記憶された指定された話者情報に対応する基準分布を、前記基準分布を推定する場合の初期値に設定することを特徴とする付記 17 に記載の音声評価方法。
- [0190] (付記 19) 前記分布を補正する処理は、予め定められた所定区間における複数のピッチ周波数を基にして、前記分布の中心部および前記分布の裾部を特定することを特徴とする付記 13 に記載の音声評価方法。
- [0191] (付記 20) 前記評価する処理は、異なる話者情報毎に音声を評価し、複数話者による会話全体を評価することを特徴とする付記 12 に記載の音声評価方法。
- [0192] (付記 21) 前記評価する処理は、各時刻における音声の印象をそれぞれ評価し、各評価結果に関するスコアを算出することを特徴とする付記 12 に記載の音声評価方法。
- [0193] (付記 22) 前記評価する処理は、前記分布の広がり大きい場合は声の印象が良いと評価し、前記分布の広がり小さい場合は声の印象が悪いと評価することを特徴とする付記 13、14 または 15 に記載の音声評価方法。
- [0194] (付記 23) 音声信号を解析してピッチ周波数を検出するピッチ検出部と、
検出した前記ピッチ周波数の検出頻度の分布に基づいて、検出した前記ピッチ周波数のうち評価対象とする評価対象領域を選択し、前記検出頻度の分布と選択した前記評価対象領域に基づいて、音声を評価する評価部と、
を有することを特徴とする音声評価装置。
- [0195] (付記 24) 前記分布の中心部に対応するピッチ周波数に第 1 の重みを乗算し、前記分布の裾部分に対応するピッチ周波数に前記第 1 の重みよりも小さい第 2 の重みを乗算することで、前記分布を補正する広がり算出部を更に有し、前記評価部は、補正した分布の広がりに基づいて、前記発話区間内の音声の印象を評価することを特徴とする付記 23 に記載の音声評価装置。
- [0196] (付記 25) 前記ピッチ検出部は、各時刻のピッチ周波数と、一定時間内に

おける平均のピッチ周波数、上限のピッチ周波数、下限のピッチ周波数のいずれかを検出することを特徴とする付記 2 3 または 2 4 に記載の音声評価装置。

[0197] (付記 2 6) 前記広がり算出部は、複数のピッチ周波数に関する、分散、標準偏差、範囲、平方和、四分位範囲のいずれかの統計量を算出し、前記評価部は、前記統計量を基にして、音声の印象を評価することを特徴とする付記 2 4 または 2 5 に記載の音声評価装置。

[0198] (付記 2 7) 予め定めた所定区間内の音声信号を解析して得られるピッチ周波数を基にして、ユーザの平常時の検出頻度の基準分布を推定する推定部を更に有し、前記評価部は、前記基準分布と、補正された分布とを基にして、音声の印象を評価することを特徴とする付記 2 4 に記載の音声評価装置。

[0199] (付記 2 8) 前記推定部は、前記基準分布と、話者情報とを対応づけて記憶装置に記憶し、前記評価部は、話者情報に対応する基準分布を選択し、選択した基準分布の広がりと、補正された分布の広がりとを基にして、音声の印象を評価することを特徴とする付記 2 7 に記載の音声評価装置。

[0200] (付記 2 9) 前記推定部は、指定された話者情報に対応する基準分布が前記記憶装置に記憶されている場合に、記憶装置に記憶された指定された話者情報に対応する基準分布を、前記基準分布を推定する場合の初期値に設定することを特徴とする付記 2 8 に記載の音声評価装置。

[0201] (付記 3 0) 前記広がり算出部は、予め定められた所定区間における複数のピッチ周波数を基にして、前記分布の中心部および前記分布の裾部を特定することを特徴とする付記 2 4 に記載の音声評価装置。

[0202] (付記 3 1) 前記評価部は、異なる話者情報毎に音声を評価し、複数話者の会話全体を評価することを特徴とする付記 2 3 に記載の音声評価装置。

[0203] (付記 3 2) 前記評価部は、異なる発話区間の音声をそれぞれ評価し、各評価結果に関するスコアを算出することを特徴とする付記 2 3 に記載の音声評価装置。

[0204] (付記 3 3) 前記評価部は、前記分布の広がりが大きい場合は声の印象が良いと評価し、前記分布の広がりが小さい場合は声の印象が悪いと評価することを特徴とする付記 2 4、2 5 または 2 6 に記載の音声評価装置。

符号の説明

[0205] 1 0 0, 2 0 0, 4 0 0 音声評価装置
1 5 0, 3 0 0 収録機器

請求の範囲

- [請求項1] 音声信号を解析してピッチ周波数を検出し、
検出した前記ピッチ周波数の検出頻度の分布に基づいて、検出した前記ピッチ周波数のうち評価対象とする評価対象領域を選択し、
前記検出頻度の分布と選択した前記評価対象領域に基づいて、音声
を評価する
処理をコンピュータに実行させることを特徴とする音声評価プログラム。
- [請求項2] 前記分布の中心部に対応するピッチ周波数に第1の重みを乗算し、
前記分布の裾部分に対応するピッチ周波数に前記第1の重みよりも小さい第2の重みを乗算することで、前記分布を補正する処理を更に実行させ、前記評価する処理は、補正した分布の広がりに基づいて、前記発話区間内の音声の印象を評価することを特徴とする請求項1に記載の音声評価プログラム。
- [請求項3] 前記ピッチ周波数を検出する処理は、各時刻のピッチ周波数と、一定時間内における平均のピッチ周波数、上限のピッチ周波数、下限のピッチ周波数のいずれかを検出することを特徴とする請求項1または2に記載の音声評価プログラム。
- [請求項4] 前記評価する処理は、複数のピッチ周波数に関する、分散、標準偏差、範囲、平方和、四分位範囲のいずれかの統計量を算出し、前記統計量を基にして、音声の印象を評価することを特徴とする請求項1、2または3に記載の音声評価プログラム。
- [請求項5] 予め定めた所定区間内の音声信号を解析して得られるピッチ周波数を基にして、ユーザの平常時の検出頻度の基準分布を推定する処理を更に実行させ、前記評価する処理は、前記基準分布と、補正された分布とを基にして、音声の印象を評価することを特徴とする請求項2に記載の音声評価プログラム。
- [請求項6] 前記基準分布と、話者情報とを対応づけて記憶装置に記憶する処理

を更に実行させ、前記評価する処理は、話者情報に対応する基準分布を選択し、選択した基準分布の広がりと、補正された分布の広がりとを基にして、音声の印象を評価することを特徴とする請求項5に記載の音声評価プログラム。

[請求項7] 前記基準分布を推定する処理は、指定された話者情報に対応する基準分布が前記記憶装置に記憶されている場合に、記憶装置に記憶された指定された話者情報に対応する基準分布を、前記基準分布を推定する場合の初期値に設定することを特徴とする請求項6に記載の音声評価プログラム。

[請求項8] 前記分布を補正する処理は、予め定められた所定区間における複数のピッチ周波数を基にして、前記分布の中心部および前記分布の裾部を特定することを特徴とする請求項2に記載の音声評価プログラム。

[請求項9] 前記評価する処理は、異なる話者情報毎に音声を評価し、複数話者による会話全体を評価することを特徴とする請求項1に記載の音声評価プログラム。

[請求項10] 前記評価する処理は、各時刻における音声の印象をそれぞれ評価し、各評価結果に関するスコアを算出することを特徴とする請求項1に記載の音声評価プログラム。

[請求項11] 前記評価する処理は、前記分布の広がりが大きい場合は声の印象が良いと評価し、前記分布の広がりが小さい場合は声の印象が悪いと評価することを特徴とする請求項2、3または4に記載の音声評価プログラム。

[請求項12] コンピュータが実行する音声評価方法であって、
音声信号を解析してピッチ周波数を検出し、
検出した前記ピッチ周波数の検出頻度の分布に基づいて、検出した前記ピッチ周波数のうち評価対象とする評価対象領域を選択し、
前記検出頻度の分布と選択した前記評価対象領域に基づいて、音声を評価する

処理をコンピュータに実行させることを特徴とする音声評価方法。

[請求項13] 前記分布の中心部に対応するピッチ周波数に第1の重みを乗算し、前記分布の裾部分に対応するピッチ周波数に前記第1の重みよりも小さい第2の重みを乗算することで、前記分布を補正する処理を更に実行させ、前記評価する処理は、補正した分布の広がりに基づいて、前記発話区間内の音声の印象を評価することを特徴とする請求項12に記載の音声評価方法。

[請求項14] 前記ピッチ周波数を検出する処理は、各時刻のピッチ周波数と、一定時間内における平均のピッチ周波数、上限のピッチ周波数、下限のピッチ周波数のいずれかを検出することを特徴とする請求項12または13に記載の音声評価方法。

[請求項15] 前記評価する処理は、複数のピッチ周波数に関する、分散、標準偏差、範囲、平方和、四分位範囲のいずれかの統計量を算出し、前記統計量に基づいて、音声の印象を評価することを特徴とする請求項12、13または14に記載の音声評価方法。

[請求項16] 予め定めた所定区間内の音声信号を解析して得られるピッチ周波数に基づいて、ユーザの平常時の検出頻度の基準分布を推定する処理を更に実行させ、前記評価する処理は、前記基準分布と、補正された分布とに基づいて、音声の印象を評価することを特徴とする請求項13に記載の音声評価方法。

[請求項17] 前記基準分布と、話者情報とを対応づけて記憶装置に記憶する処理を更に実行させ、前記評価する処理は、話者情報に対応する基準分布を選択し、選択した基準分布の広がり、補正された分布の広がりに基づいて、音声の印象を評価することを特徴とする請求項16に記載の音声評価方法。

[請求項18] 前記基準分布を推定する処理は、指定された話者情報に対応する基準分布が前記記憶装置に記憶されている場合に、記憶装置に記憶された指定された話者情報に対応する基準分布を、前記基準分布を推定す

る場合の初期値に設定することを特徴とする請求項 17 に記載の音声評価方法。

[請求項19] 前記分布を補正する処理は、予め定められた所定区間における複数のピッチ周波数を基にして、前記分布の中心部および前記分布の裾部を特定することを特徴とする請求項 13 に記載の音声評価方法。

[請求項20] 前記評価する処理は、異なる話者情報毎に音声を評価し、複数話者による会話全体を評価することを特徴とする請求項 12 に記載の音声評価方法。

[請求項21] 前記評価する処理は、各時刻における音声の印象をそれぞれ評価し、各評価結果に関するスコアを算出することを特徴とする請求項 12 に記載の音声評価方法。

[請求項22] 前記評価する処理は、前記分布の広がり大きい場合は声の印象が良いと評価し、前記分布の広がり小さい場合は声の印象が悪いと評価することを特徴とする請求項 13、14 または 15 に記載の音声評価方法。

[請求項23] 音声信号を解析してピッチ周波数を検出するピッチ検出部と、
検出した前記ピッチ周波数の検出頻度の分布に基づいて、検出した前記ピッチ周波数のうち評価対象とする評価対象領域を選択し、前記検出頻度の分布と選択した前記評価対象領域に基づいて、音声を評価する評価部と、
を有することを特徴とする音声評価装置。

[請求項24] 前記分布の中心部に対応するピッチ周波数に第 1 の重みを乗算し、前記分布の裾部分に対応するピッチ周波数に前記第 1 の重みよりも小さい第 2 の重みを乗算することで、前記分布を補正する広がり算出部を更に有し、前記評価部は、補正した分布の広がりに基づいて、前記発話区間内の音声の印象を評価することを特徴とする請求項 23 に記載の音声評価装置。

[請求項25] 前記ピッチ検出部は、各時刻のピッチ周波数と、一定時間内におけ

る平均のピッチ周波数、上限のピッチ周波数、下限のピッチ周波数のいずれかを検出することを特徴とする請求項 23 または 24 に記載の音声評価装置。

[請求項26] 前記広がり算出部は、複数のピッチ周波数に関する、分散、標準偏差、範囲、平方和、四分位範囲のいずれかの統計量を算出し、前記評価部は、前記統計量を基にして、音声の印象を評価することを特徴とする請求項 24 または 25 に記載の音声評価装置。

[請求項27] 予め定めた所定区間内の音声信号を解析して得られるピッチ周波数を基にして、ユーザの平常時の検出頻度の基準分布を推定する推定部を更に有し、前記評価部は、前記基準分布と、補正された分布とを基にして、音声の印象を評価することを特徴とする請求項 24 に記載の音声評価装置。

[請求項28] 前記推定部は、前記基準分布と、話者情報とを対応づけて記憶装置に記憶し、前記評価部は、話者情報に対応する基準分布を選択し、選択した基準分布の広がりと、補正された分布の広がりとを基にして、音声の印象を評価することを特徴とする請求項 27 に記載の音声評価装置。

[請求項29] 前記推定部は、指定された話者情報に対応する基準分布が前記記憶装置に記憶されている場合に、記憶装置に記憶された指定された話者情報に対応する基準分布を、前記基準分布を推定する場合の初期値に設定することを特徴とする請求項 28 に記載の音声評価装置。

[請求項30] 前記広がり算出部は、予め定められた所定区間における複数のピッチ周波数を基にして、前記分布の中心部および前記分布の裾部を特定することを特徴とする請求項 24 に記載の音声評価装置。

[請求項31] 前記評価部は、異なる話者情報毎に音声を評価し、複数話者の会話全体を評価することを特徴とする請求項 23 に記載の音声評価装置。

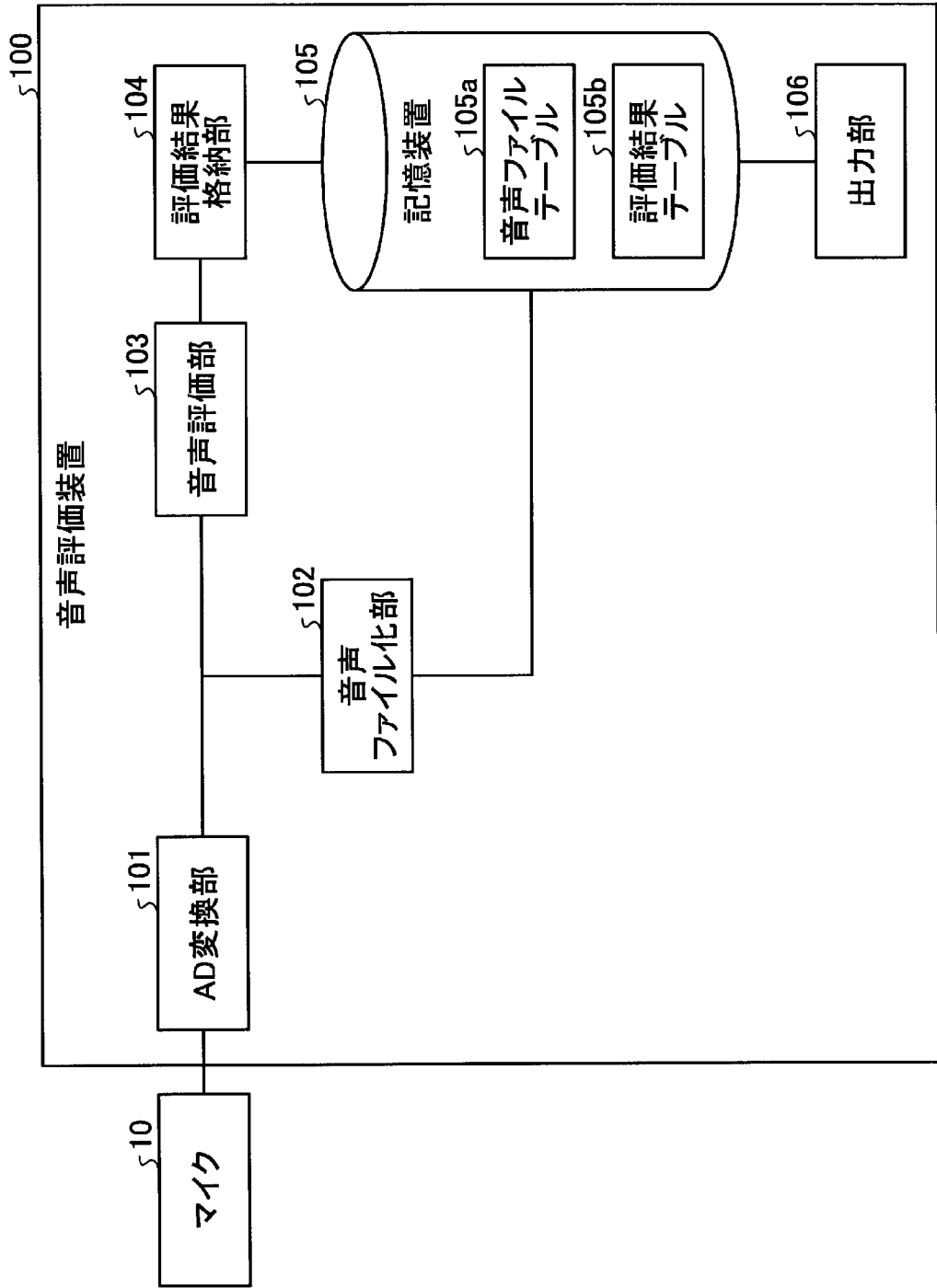
[請求項32] 前記評価部は、異なる発話区間の音声をそれぞれ評価し、各評価結

果に関するスコアを算出することを特徴とする請求項 23 に記載の音声評価装置。

[請求項33] 前記評価部は、前記分布の広がり大きい場合は声の印象が良いと評価し、前記分布の広がり小さい場合は声の印象が悪いと評価することを特徴とする請求項 24、25 または 26 に記載の音声評価装置。

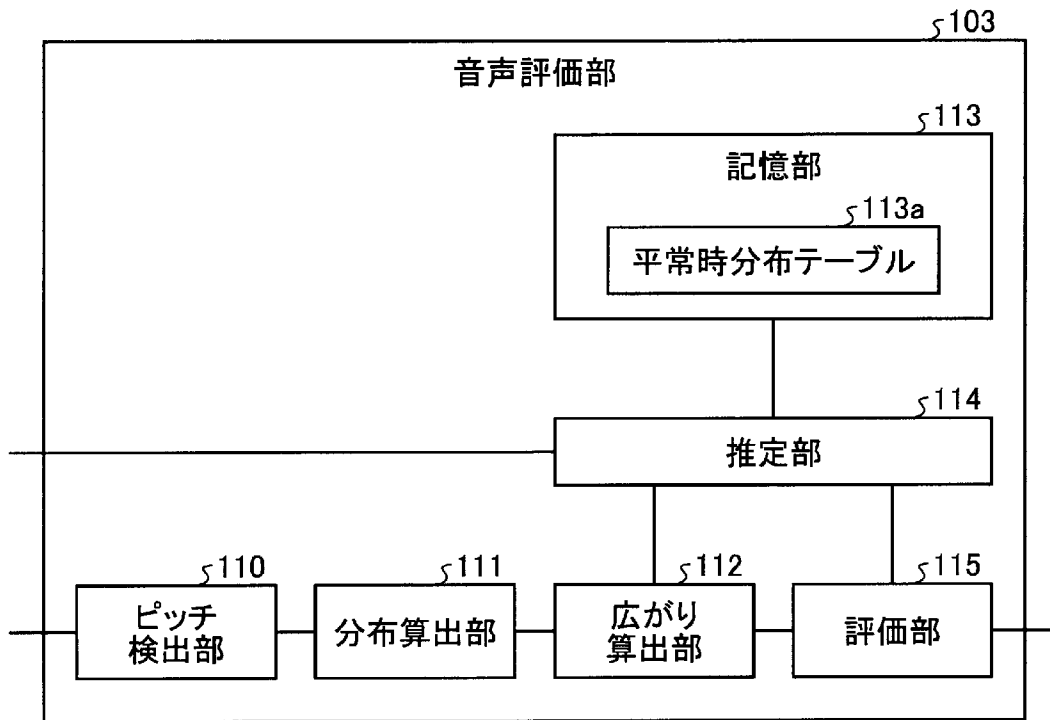
[図1]

本実施例1に係る音声評価装置の構成を示す機能ブロック図



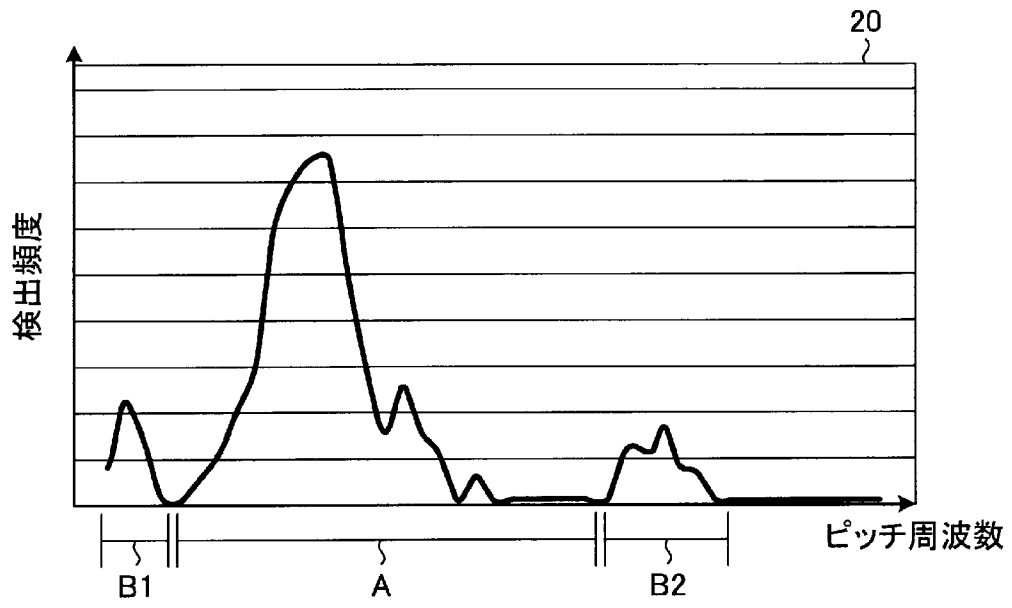
[図2]

本実施例1に係る音声評価部の構成を示す機能ブロック図



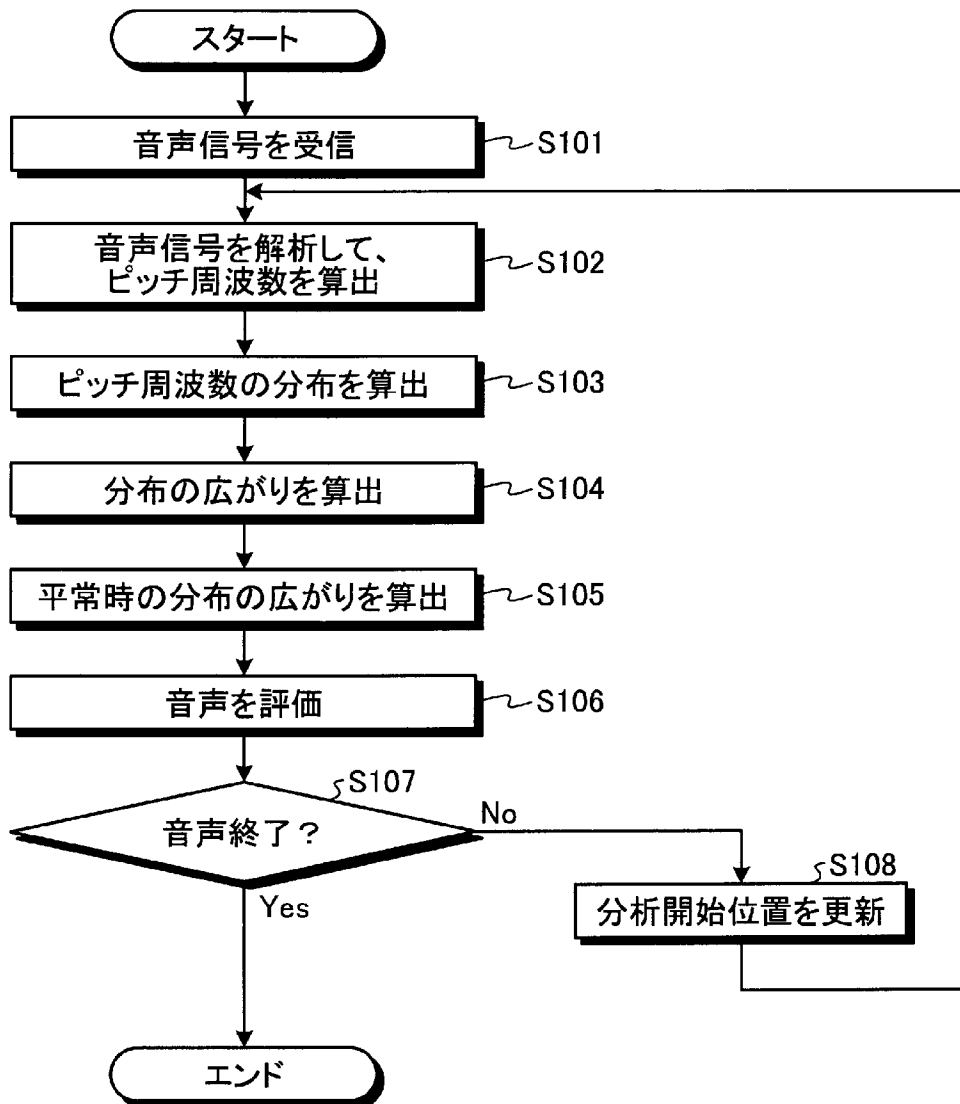
[図3]

ヒストグラムの一例を示す図



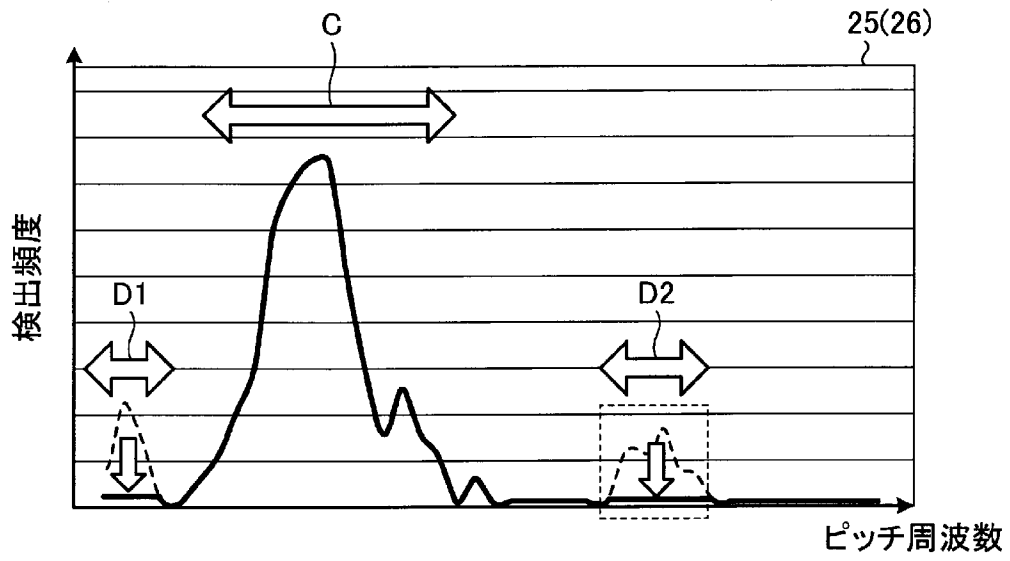
[図4]

本実施例1に係る音声評価部の処理手順を示すフローチャート



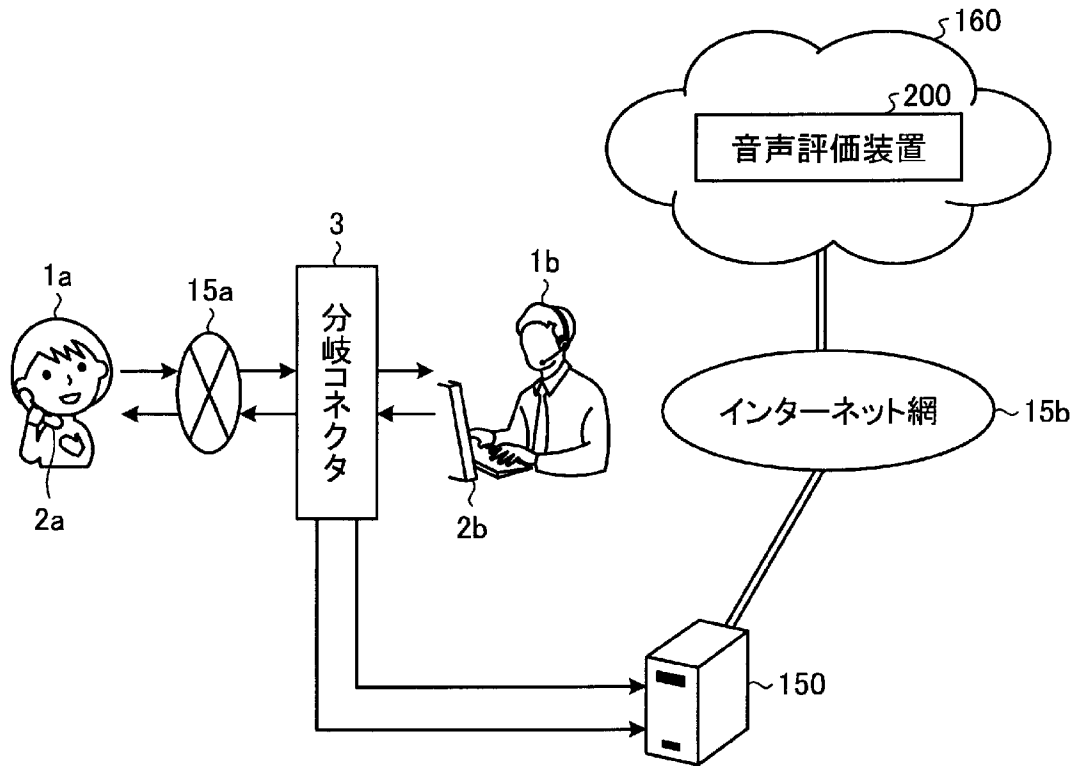
[図5]

音声評価装置のその他の処理を説明するための図



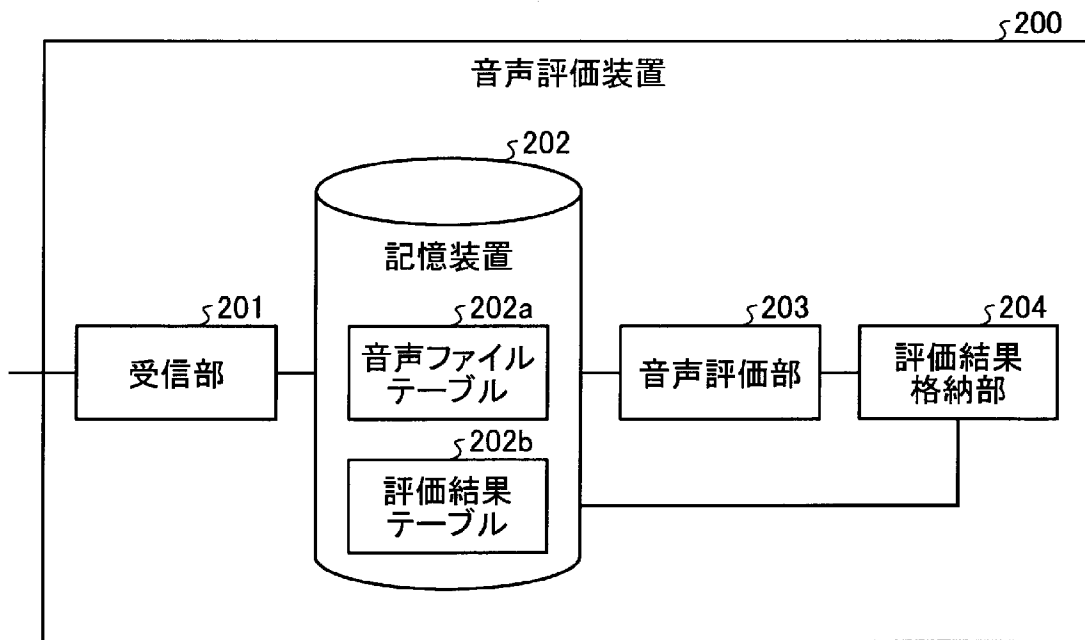
[図6]

本実施例2に係る音声評価システムの構成を示す図



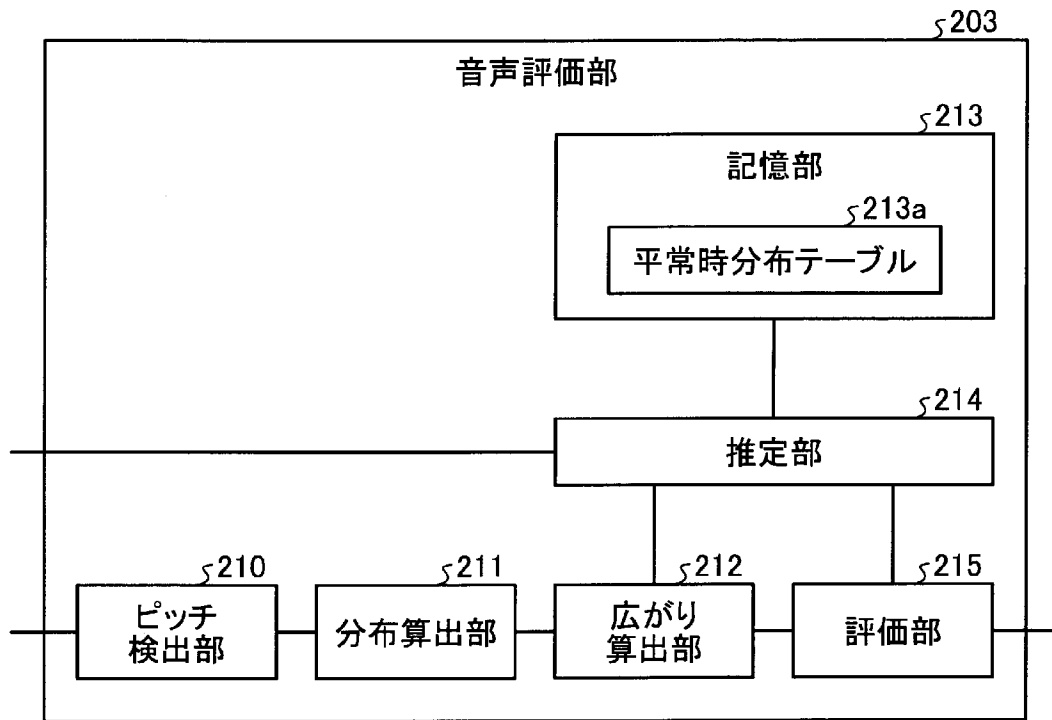
[図7]

本実施例2に係る音声評価装置の構成を示す機能ブロック図



[図8]

本実施例2に係る音声評価部の構成を示す機能ブロック図



[図9]

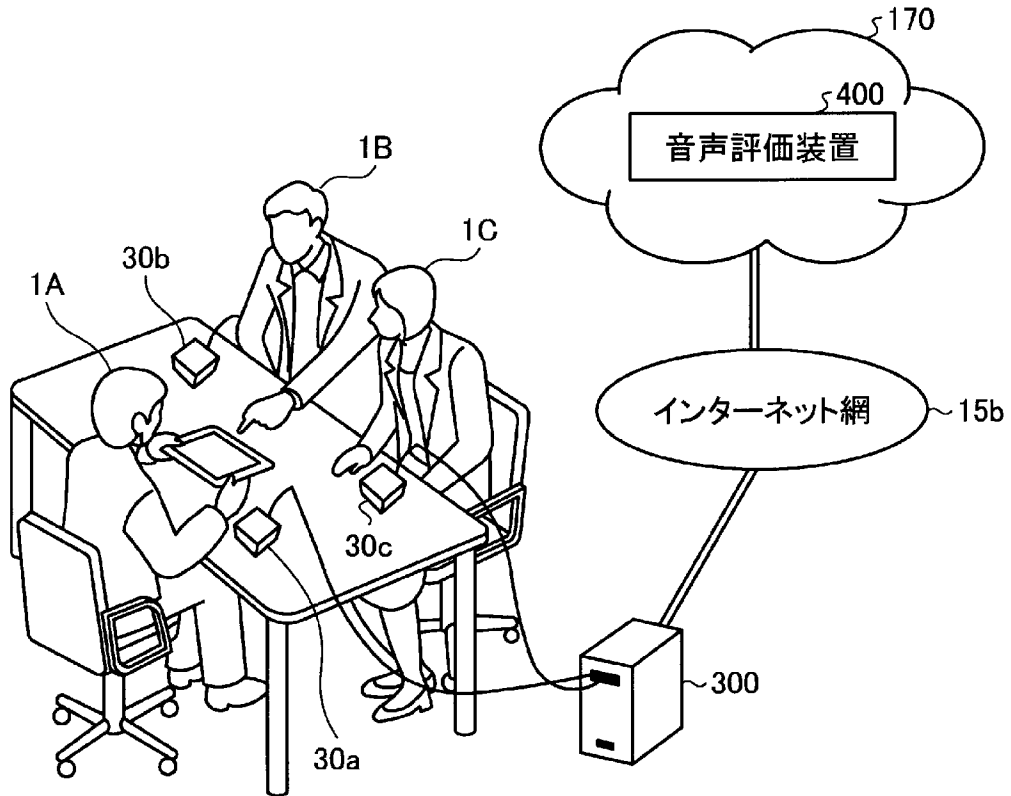
本実施例2に係る平常時分布テーブルのデータ構造の一例を示す図

5213a

属性情報		標準偏差 (ヒストグラムの広がり)
性別情報	声の高さ情報	
男性	低い	性別情報「男性」、声の高さ情報「低い」に対応する平常時の標準偏差
男性	高い	性別情報「男性」、声の高さ情報「高い」に対応する平常時の標準偏差
女性	低い	性別情報「女性」、声の高さ情報「低い」に対応する平常時の標準偏差
女性	高い	性別情報「女性」、声の高さ情報「高い」に対応する平常時の標準偏差

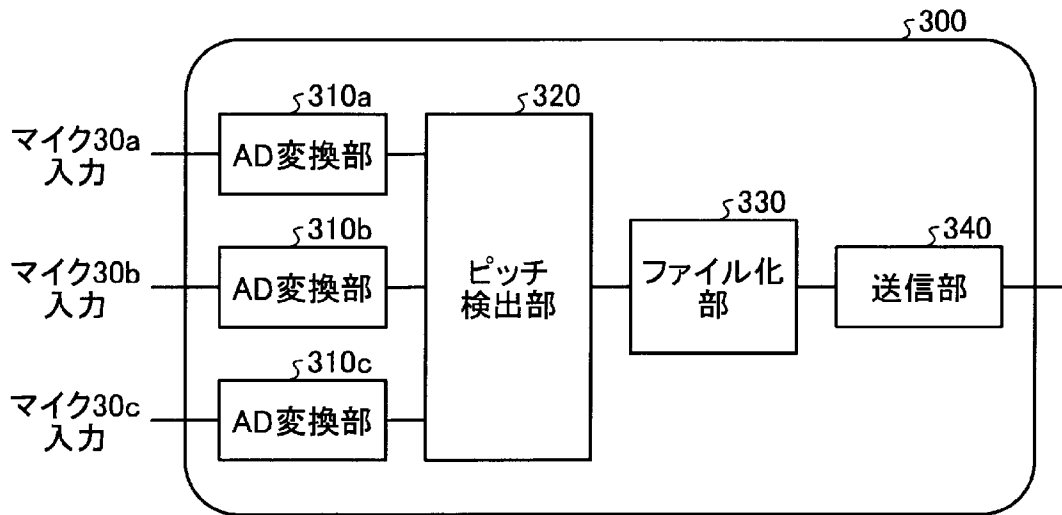
[図10]

本実施例3に係る音声評価システムの構成を示す図



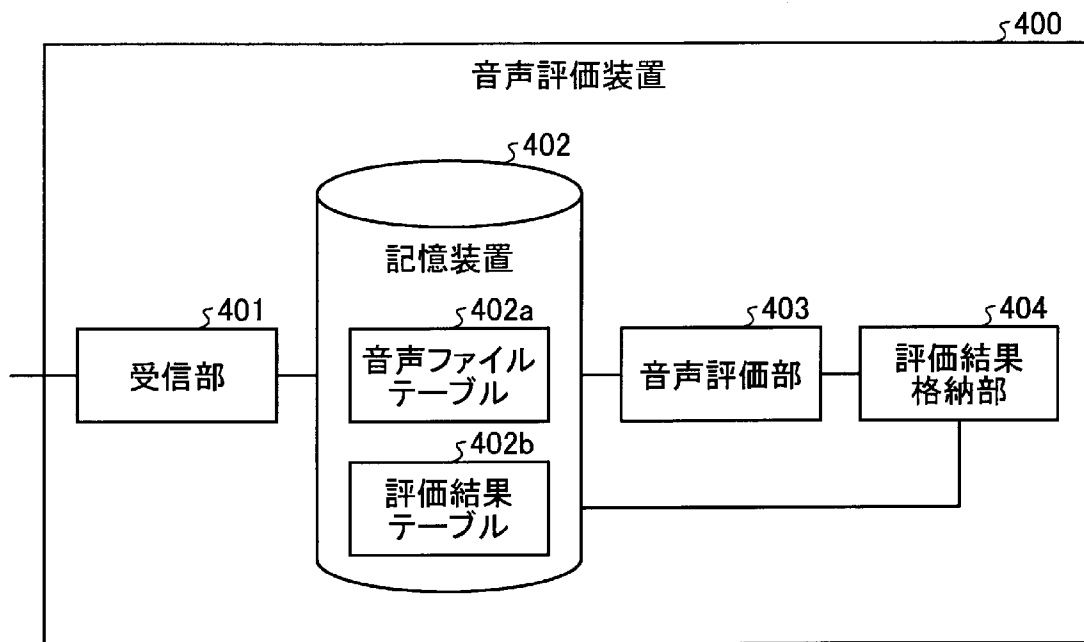
[図11]

本実施例3に係る収録機器の構成を示す機能ブロック図



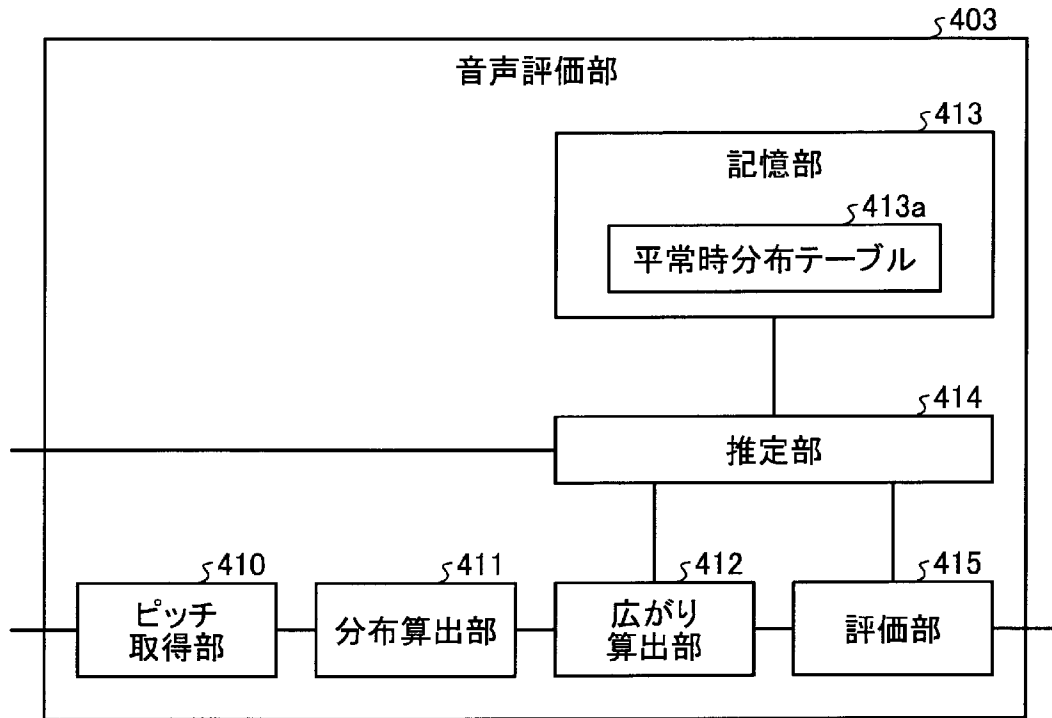
[図12]

本実施例3に係る音声評価装置の構成を示す機能ブロック図



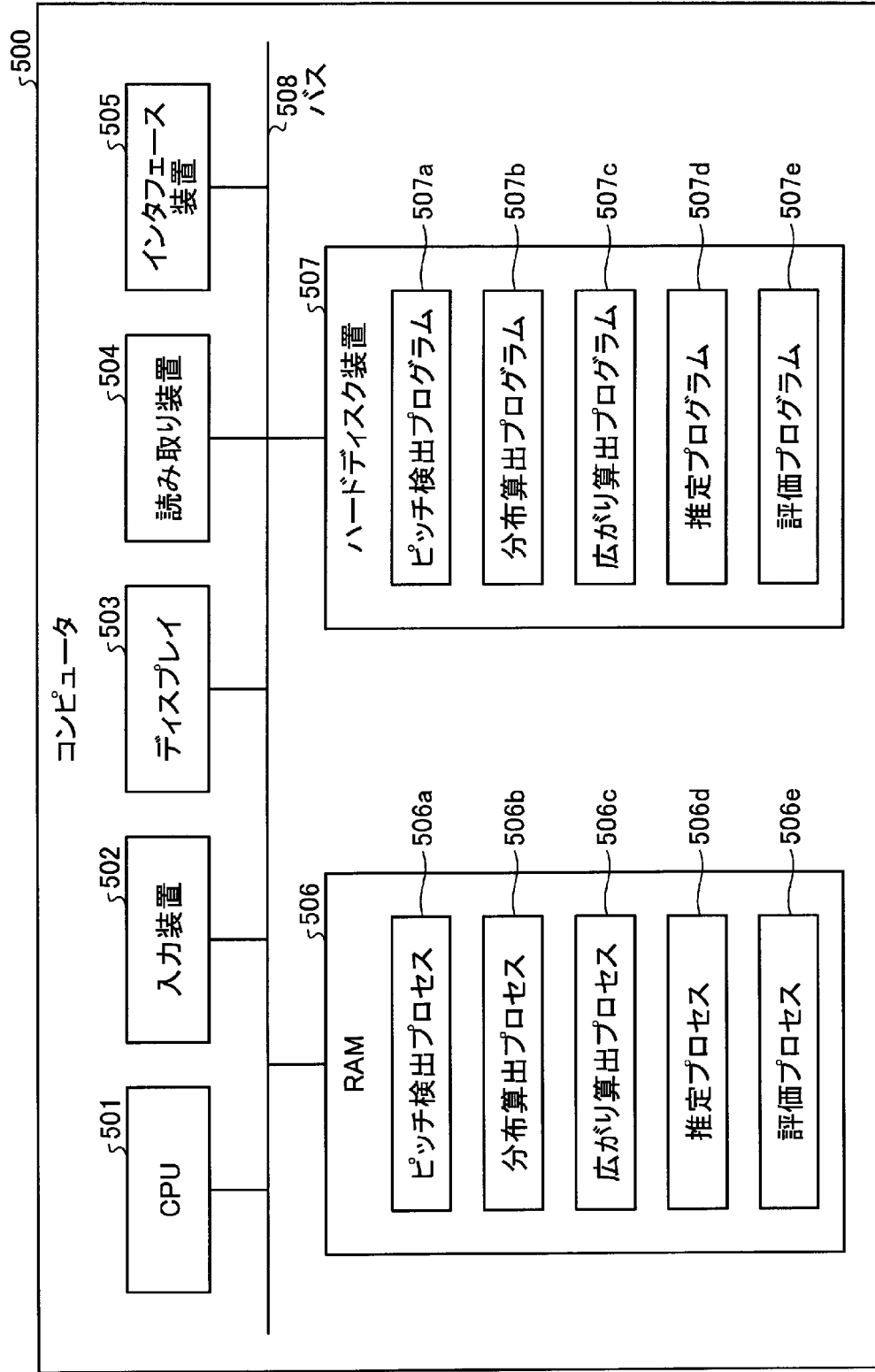
[図13]

本実施例3に係る音声評価部の構成を示す機能ブロック図



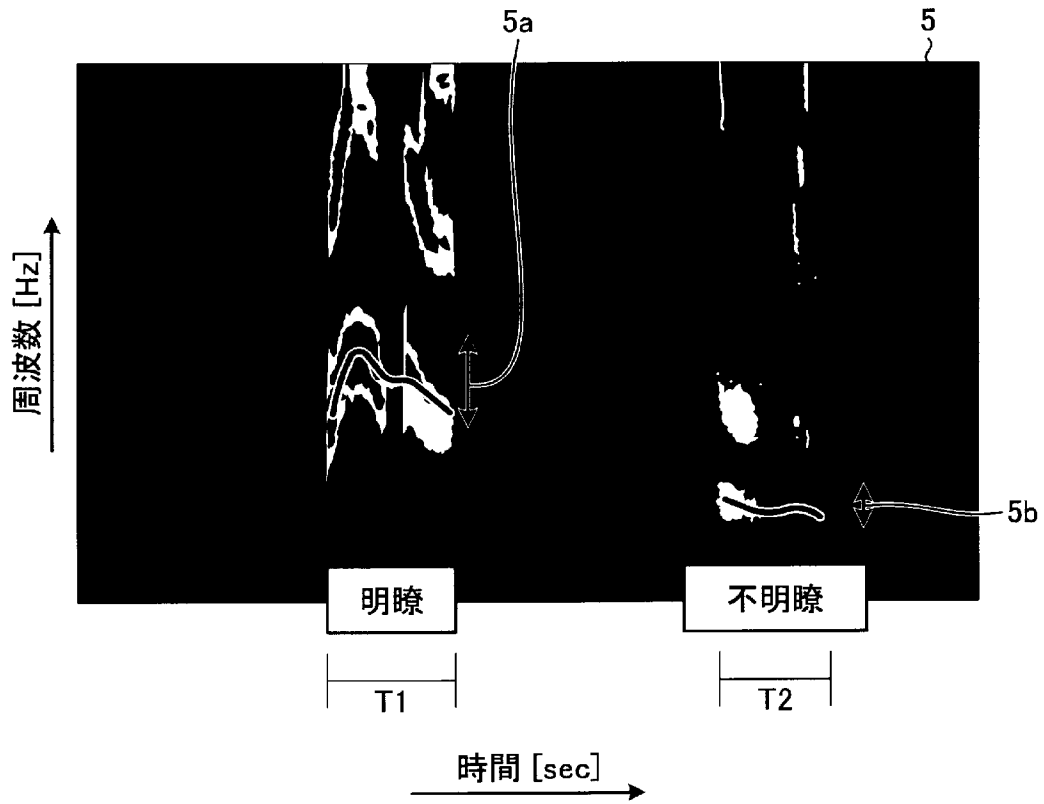
[図14]

音声評価装置と同様の機能を実現するコンピュータのハードウェア構成の一例を示す図



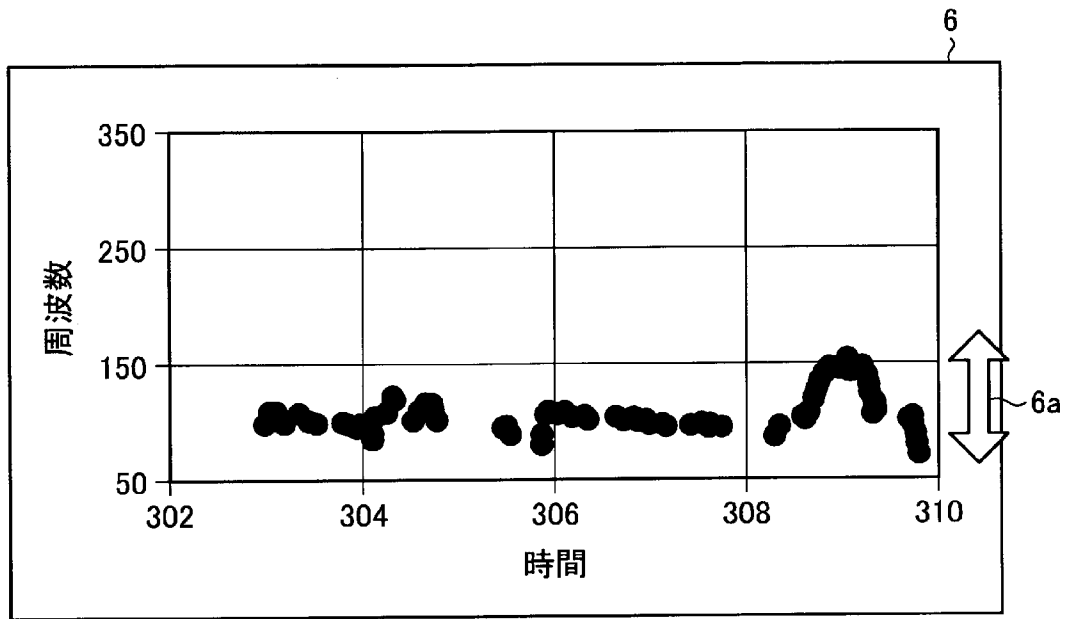
[図15]

従来技術を説明するための図



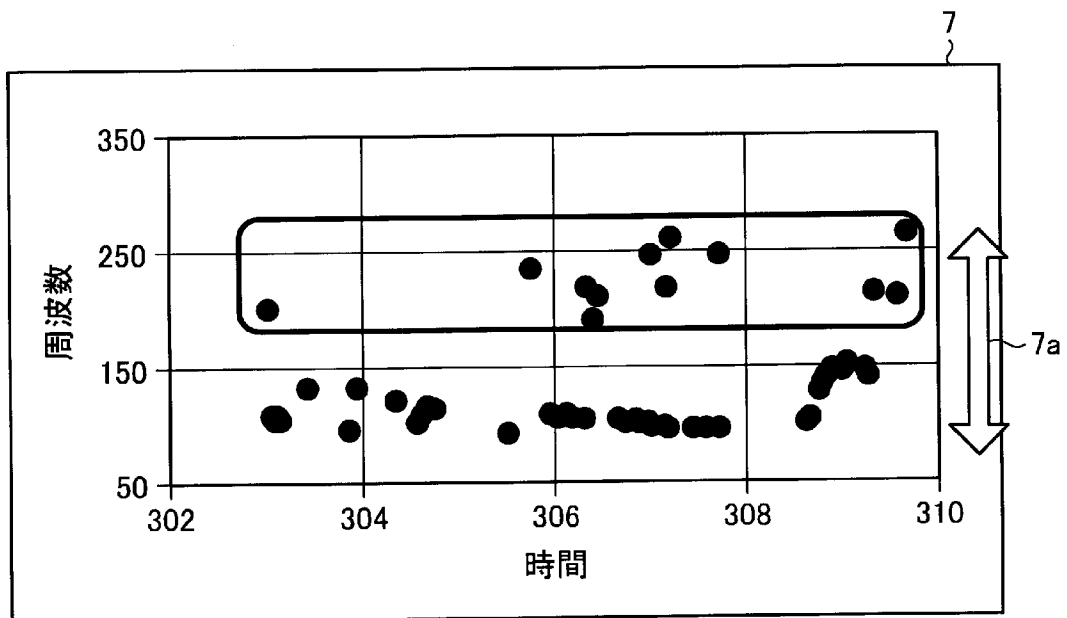
[図16]

正確なピッチ周波数を算出した場合のグラフを示す図



[図17]

半ピッチ/倍ピッチを算出した場合のグラフを示す図



INTERNATIONAL SEARCH REPORT

International application No. PCT/JP2018/013867
--

A. CLASSIFICATION OF SUBJECT MATTER
 Int.Cl. G10L25/60 (2013.01) i, G10L25/90 (2013.01) 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)
 Int.Cl. G10L25/60, G10L25/90

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

Published examined utility model applications of Japan	1922-1996
Published unexamined utility model applications of Japan	1971-2018
Registered utility model specifications of Japan	1996-2018
Published registered utility model applications of Japan	1994-2018

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2006-267464 A (TOKYO ELECTRIC POWER CO.; NIPPON ONKYO KENKYUSHO KK) 05 October 2006, paragraph [0031], fig. 8 (Family: none)	1-33
A	JP 2006-267465 A (TOKYO ELECTRIC POWER CO.; NIPPON ONKYO KENKYUSHO KK) 05 October 2006, paragraphs [0037]-[0039], fig. 11 (Family: none)	1-33
A	JP 2013-72979 A (FUJI XEROX CO., LTD.) 22 April 2013, paragraphs [0038]-[0040] & US 2013/0080169 A1, paragraphs [0058]-[0065]	1-33

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	“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
“E” earlier application or patent but published on or after the international filing date	“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
“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)	“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
“O” document referring to an oral disclosure, use, exhibition or other means	“&” document member of the same patent family
“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 05.06.2018	Date of mailing of the international search report 12.06.2018
---	--

Name and mailing address of the ISA/ Japan Patent Office 3-4-3, Kasumigaseki, Chiyoda-ku, Tokyo 100-8915, Japan	Authorized officer Telephone No.
--	---

A. 発明の属する分野の分類（国際特許分類（IPC）） Int.Cl. G10L25/60(2013.01)i, G10L25/90(2013.01)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. G10L25/60, G10L25/90			
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 1971-2018年 日本国実用新案登録公報 1996-2018年 日本国登録実用新案公報 1994-2018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）			
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	
A	JP 2006-267464 A（東京電力株式会社，有限会社日本音響研究所） 2006.10.05，段落 [0031]，図8（ファミリーなし）	1-33	
A	JP 2006-267465 A（東京電力株式会社，有限会社日本音響研究所） 2006.10.05，段落 [0037] - [0039]，図11（ファミリーなし）	1-33	
A	JP 2013-72979 A（富士ゼロックス株式会社）2013.04.22，段落 [0038] - [0040] & US 2013/0080169 A1, [0058]-[0065]	1-33	
☐ 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	05.06.2018	国際調査報告の発送日	12.06.2018
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） 上田 雄 電話番号 03-3581-1101 内線 3591	5Z 5095