

(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9月27日(27.09.2018)



(10) 国際公開番号

WO 2018/173335 A1

(51) 国際特許分類:

G06F 17/30 (2006.01)

(21) 国際出願番号: PCT/JP2017/034573

(22) 国際出願日: 2017年9月25日(25.09.2017)

(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 2017-056495 2017年3月22日(22.03.2017) JP

(71) 出願人: 株式会社東芝 (KABUSHIKI KAISHA TOSHIBA) [JP/JP]; 〒1058001 東京都港区芝浦一丁目1番1号 Tokyo (JP). 東芝デジタルソリューションズ株式会社 (TOSHIBA DIGITAL SOLUTIONS CORPORATION) [JP/JP]; 〒2128585 神奈川県川崎市幸区堀川町72番地34 Kanagawa (JP).

(72) 発明者: 森 紘一郎 (MORI, Kouichirou); 〒1058001 東京都港区芝浦一丁目1番1号 株式会社東芝 知的財産室内 Tokyo (JP). 鈴木 優

(SUZUKI, Masaru); 〒2128585 神奈川県川崎市幸区堀川町72番地34 東芝デジタルソリューションズ株式会社内 Kanagawa (JP). 大谷 大和 (OHTANI, Yamato); 〒1058001 東京都港区芝浦一丁目1番1号 株式会社東芝 知的財産室内 Tokyo (JP). 森田 真弘 (MORITA, Masahiro); 〒2128585 神奈川県川崎市幸区堀川町72番地34 東芝デジタルソリューションズ株式会社内 Kanagawa (JP).

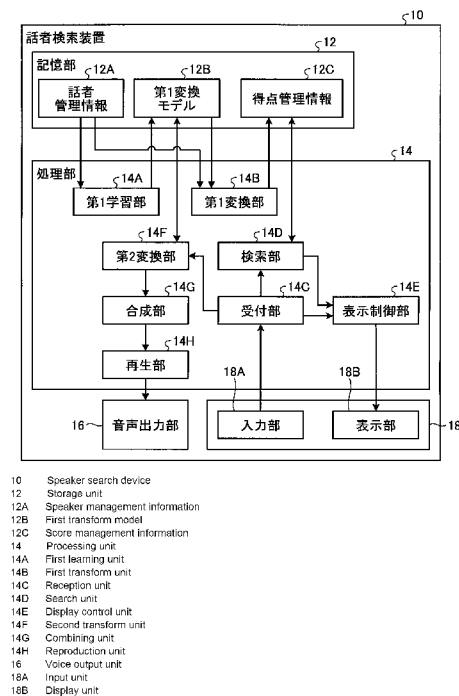
(74) 代理人: 特許業務法人酒井国際特許事務所 (SAKAI INTERNATIONAL PATENT OFFICE); 〒1000013 東京都千代田区霞が関3丁目8番1号 虎の門三井ビルディング Tokyo (JP).

(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, 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, HN, HR, HU, ID, IL, IN, IR, IS, JO, KE, KG, KH,

(54) Title: SPEAKER SEARCH DEVICE, SPEAKER SEARCH METHOD, AND SPEAKER SEARCH PROGRAM

(54) 発明の名称: 話者検索装置、話者検索方法、および話者検索プログラム

[図1]



(57) Abstract: A speaker search device (10) is provided with a first transform unit (14B), a reception unit (14C), and a search unit (14D). By using an inverse transform model of a first transform model (12B) for transforming a score vector representing a first features of voice quality into an acoustic model, the first transform unit (14B) transforms an acoustic model registered in advance into a score vector, and registers the score vector in score management information (12C) in association with a speaker ID. The reception unit (14C) receives an input of the score vector. The search unit (14D) searches the score management information (12C) for a speaker ID similar to the received score vector.

(57) 要約: 話者検索装置 (10) は、第1変換部 (14B) と、受付部 (14C) と、検索部 (14D) と、を備える。第1変換部 (14B) は、声質の特徴を表す得点ベクトルを音響モデルに変換するための第1変換モデル (12B) の逆変換モデルを用いて、予め登録された音響モデルを得点ベクトルに変換し、話者IDに対応づけて得点管理情報 (12C) に登録する。受付部 (14C) は、得点ベクトルの入力を受付ける。検索部 (14D) は、得点管理情報 (12C) から、受付けた得点ベクトルに類似する話者IDを検索する。

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))

明 細 書

発明の名称：

話者検索装置、話者検索方法、および話者検索プログラム

技術分野

[0001] 本発明の実施の形態は、話者検索装置、話者検索方法、および話者検索プログラムに関する。

背景技術

[0002] 近年、音声合成技術の発展により、テキストから高品質な合成音が作成されている。特に隠れマルコフモデル（HMM）を用いた音声合成技術を用いることで、音響モデルを操作することで合成音を柔軟に制御できることが知られている。また、音声から音響特徴量を抽出し、入力音声の音響特徴量と音声データベースの各話者の音響特徴量の類似度を求め、入力音声と声質の類似度が高い話者を検索する技術が提案されている。この技術を使えば所望の話者の音声から音響特徴量を抽出し、音響特徴量間の類似度を求めることで大規模な音声データベースから所望の話者の声質に近い話者候補を検索できる。

[0003] しかし、従来手法では、ユーザが所望する話者の音声をクエリとして事前に用意する必要があった。このため、従来では、所望の話者の音声を用意できなければ類似話者を検索できないという課題があった。すなわち、従来では、声質類似度の高い話者を検索することは困難であった。

先行技術文献

特許文献

[0004] 特許文献1：WO 2009／110613号公報

非特許文献

[0005] 非特許文献1：Yusuke Ijima et al. Similar Speaker Selection Technique Based on Distance Metric Learning Using Highly Correlated Acoustic Features with Perceptual Voice Quality Similarity

, IEICE Trans. INF. & SYST., Vol.E98-D, No.1 (2015)

発明の概要

発明が解決しようとする課題

[0006] 本発明が解決しようとする課題は、声質類似度の高い話者を検索することができる話者検索装置、話者検索方法、および話者検索プログラムを提供することである。

課題を解決するための手段

[0007] 実施の形態の話者検索装置は、第1変換部と、受付部と、検索部と、を備える。第1変換部は、声質の特徴を表す得点ベクトルを音響モデルに変換するための第1変換モデルの逆変換モデルを用いて、予め登録された音響モデルを得点ベクトルに変換し、話者識別情報に対応づけて得点管理情報に登録する。受付部は、得点ベクトルの入力を受付ける。検索部は、得点管理情報から、受付けた得点ベクトルに類似する話者識別情報を検索する。

図面の簡単な説明

- [0008] [図1]図1は、話者検索装置の一例を示す図である。
- [図2]図2は、話者管理情報のデータ構成の一例を示す模式図である。
- [図3]図3は、得点ベクトルの一例を示す模式図である。
- [図4]図4は、得点管理情報のデータ構成の一例を示す模式図である。
- [図5]図5は、入力画面の一例を示す模式図である。
- [図6]図6は、表示画面の一例を示す模式図である。
- [図7]図7は、検索処理の手順の一例を示すフローチャートである。
- [図8]図8は、話者検索処理の手順の一例を示すフローチャートである。
- [図9]図9は、話者検索装置の一例を示す図である。
- [図10]図10は、音声管理情報のデータ構成の一例を示す模式図である。
- [図11]図11は、得点管理情報のデータ構成の一例を示す模式図である。
- [図12]図12は、話者検索装置が実行する検索処理の手順の一例を示すフローチャートである。

[図13]図 1 3 は、話者検索装置のハードウェア構成例を示す説明図である。

発明を実施するための形態

[0009] 以下に添付図面を参照して、話者検索装置、話者検索方法、および話者検索プログラムを詳細に説明する。

[0010] (第 1 の実施の形態)

図 1 は、本実施の形態の話者検索装置 1 0 の一例を示す図である。

[0011] 話者検索装置 1 0 は、ユーザの所望の声質の話者を検索する装置である。話者検索装置 1 0 は、記憶部 1 2 と、処理部 1 4 と、音声出力部 1 6 と、U I (ユーザ・インターフェース) 部 1 8 と、を備える。

[0012] 記憶部 1 2、音声出力部 1 6、および U I 部 1 8 と、処理部 1 4 と、は、バスなどを介してデータや信号を授受可能に接続されている。

[0013] 音声出力部 1 6 は、音声を出力する。音声出力部 1 6 は、例えば、スピーカである。

[0014] U I 部 1 8 は、入力部 1 8 A と、表示部 1 8 B と、を含む。入力部 1 8 A は、ユーザの操作を受付ける。入力部 1 8 A は、例えば、デジタルペン、マウス、またはトラックボール等のポインティングデバイスや、キーボード等の入力デバイスである。表示部 1 8 B は、各種情報を表示する。表示部 1 8 B は、例えば、LCD などのディスプレイや、投影装置などである。なお、U I 部 1 8 は、入力部 1 8 A と表示部 1 8 B とを一体的に備えた、タッチパネルであってもよい。

[0015] 記憶部 1 2 は、各種データを記憶する。記憶部 1 2 は、例えば、RAM (Random Access Memory)、フラッシュメモリ等の半導体メモリ素子、ハードディスク、光ディスク等である。なお、記憶部 1 2 は、話者検索装置 1 0 の外部に設けられた記憶装置であってもよい。また、記憶部 1 2 は、記憶媒体であってもよい。具体的には、記憶媒体は、プログラムや各種情報を、LAN (Local Area Network) やインターネットなどを介してダウンロードして記憶または一時記憶したものであってもよい。また、記憶部 1 2 を、複数の記憶媒体から構成してもよい。

- [0016] 本実施の形態では、記憶部 12 は、話者管理情報 12 A と、第 1 変換モデル 12 B と、得点管理情報 12 C と、を記憶する。
- [0017] 話者管理情報 12 A は、話者の音声に関する情報である。話者管理情報 12 A は、記憶部 12 に予め記憶されている。図 2 は、話者管理情報 12 A のデータ構成の一例を示す模式図である。
- [0018] 話者管理情報 12 A は、話者 I D と、音声データと、音声のテキストと、音響特徴量と、言語特徴量と、音響モデルと、得点ベクトルと、を対応づけたデータベースである。なお、話者管理情報 12 A のデータ形式は、データベースに限定されない。例えば、話者管理情報 12 A は、テーブルであってもよい。
- [0019] 話者 I D は、話者を識別する識別情報（話者識別情報）である。話者 I D は、話者を識別可能な情報であればよい。話者 I D は、例えば、話者の名前などである。
- [0020] 音声データは、対応する話者 I D によって識別される話者の音声のデータである。話者管理情報 12 A には、話者 I D に対応づけて、複数の音声データ（例えば N 個（N は 2 以上の整数））が登録されている。音声データのデータフォーマットは限定されない。例えば、音声データのデータフォーマットは、WAV、AIFF、BWF などである。
- [0021] 音声のテキストは、音声データの収録原稿である。
- [0022] 音響特徴量は、対応する音声データから抽出された、音声（音響）の特徴量である。音響特徴量には、HMM（隠れマルコフモデル（hidden Markov model））音声合成で使われる音響特徴量を用いればよい。例えば、音響特徴量は、音韻や声色を表すメルケプストラム係数、メル LPC 係数、メル LSP 係数、声の高さを表す基本周波数（F0）、音声の周期・非周期成分の割合を表す非周期性指標（BAP）などである。話者管理情報 12 A には、登録されている話者 I D に対応する音響特徴量が予め導出され、予め話者管理情報 12 A に登録されているものとする。
- [0023] 言語特徴量は、対応する音声のテキストから抽出された、言語の特徴量で

ある。例えば、言語特徴量は、前後の音素、発音に関する情報、句末位置、文長、アクセント句長、モーラ長、モーラ位置、アクセント型、品詞、係り受け情報などである。話者管理情報 12A には、登録されている話者 ID に対応する言語特徴量が予め導出され、予め話者管理情報 12A に登録されているものとする。

[0024] 音響モデルは、話者の音声の言語特徴量から音声の音響特徴量への写像を統計的にモデル化したものである。例えば、音響モデルには、HMM 音声合成における音響モデルを用いる。音響モデルは、入力を言語特徴量、出力を音響特徴量、とした決定木でモデル化される。出力である音響特徴量は、決定木の各リーフノードに割り当てられた多次元正規分布の平均ベクトルと共分散行列の集合によって、表現される。話者の音響特徴量と言語特徴量から、音響モデルのパラメータである上記平均ベクトルと共分散行列を推定することによって、音響モデルが生成される。

[0025] 本実施の形態では、音響モデルとして、HMM 音声合成における決定木の音響モデルを用いる場合を説明する。しかし、音響モデルとして、HMM 音声合成における決定木の音響モデルを用いる形態に限定されない。例えば、ニューラルネットワークなどの言語特徴量を音響特徴量に写像する任意のモデルを、音響モデルとして用いてもよい。

[0026] 話者管理情報 12A には、登録されている話者 ID に対応する音響モデルが予め導出され、予め話者管理情報 12A に予め登録されているものとする。なお、本実施の形態では、話者管理情報 12A における、一部の話者 ID には、音響モデルが対応付けて登録されていなくてもよい。

[0027] 得点ベクトルは、話者の声質の特徴を表す。本実施の形態では、得点ベクトルは、話者の声質の主観的な特徴を表す。

[0028] 図 3 は、得点ベクトルの一例を示す模式図である。得点ベクトルは、話者の声質の特徴の主観的な強さを、声質の特徴の種類ごとに得点（すなわち数値）で表したものである。声質の特徴の種類は、物理特徴量ではなく、声質を主観的に分類したものである。声質の特徴の種類は、予め定めればよい。

例えば、声質の特徴の種類は、「性別」、「年齢」、「明るさ」、「硬さ」、「明瞭さ」、「流暢さ」、「かすれ」、などである。すなわち、得点ベクトルは、話者の主観的な特徴を、声質を表す主観的な種類に応じたラベルの付与された軸を用いて表した、ベクトルである。

[0029] これらの特徴の種類に対応する得点は、これらの種類の特徴の強さを表す。得点は、例えば、数値で表される。図3には、得点を、“0”から“1”の範囲で表した場合を示している。例えば、声質の特徴の種類「性別」は、得点“0”に近いほど女性的であり、得点“1”に近いほど男性的であることを示す。また、声質の特徴の種類「年齢」は、得点“0”に近いほど若く、得点“1”に近いほど年齢が高いことを示す。

[0030] このため、図3に示す例では、話者ID“A”に対応する得点ベクトルは、該話者ID“A”によって識別される話者の声質の特徴が、若い子供の声であることを示す。また、話者ID“B”に対応する得点ベクトルは、該話者ID“B”によって識別される話者の声質の特徴が、若い女性の声であることを示す。また、話者ID“D”に対応する得点ベクトルは、該話者ID“D”によって識別される話者の声質の特徴が、年老いた男性の声であることを示す。

[0031] 図2に戻り説明を続ける。話者管理情報12Aには、話者管理情報12Aに登録されている複数の話者IDの内、一部の話者IDに対応づけて、得点ベクトルが予め登録されているものとする。話者管理情報12Aの一部の話者IDには、該話者IDによって識別される話者の声質に関するアンケート調査によって、多くの人の主観的な声質の感じ方が反映されるように、予め得点ベクトルが付与されているものとする。

[0032] なお、話者管理情報12Aに登録されている話者IDの数は、ユーザが所望の話者を話者管理情報12Aから手作業で検索することが困難な数以上であるものとする。

[0033] 図1に戻り、説明を続ける。次に、第1変換モデル12Bおよび得点管理情報12Cについて説明する。第1変換モデル12Bは、得点ベクトルを音

響モデルに変換するためのモデルである。第1変換モデル12Bは、処理部14によって学習され、記憶部12へ記憶される（詳細後述）。

[0034] 得点管理情報12Cは、ユーザによる話者の検索時に用いられるデータベースである。得点管理情報12Cは、後述する処理部14によって登録される（詳細後述）。

[0035] 次に、処理部14について説明する。処理部14は、話者検索装置10を制御する。処理部14は、第1学習部14Aと、第1変換部14Bと、受付部14Cと、検索部14Dと、表示制御部14Eと、第2変換部14Fと、合成部14Gと、再生部14Hと、を含む。

[0036] 上記各部（第1学習部14A、第1変換部14B、受付部14C、検索部14D、表示制御部14E、第2変換部14F、合成部14G、再生部14H）は、例えば、1または複数のプロセッサにより実現される。例えば上記各部は、CPU（Central Processing Unit）などのプロセッサにプログラムを実行させること、すなわちソフトウェアにより実現してもよい。上記各部は、専用のIC（Integrated Circuit）などのプロセッサ、すなわちハードウェアにより実現してもよい。上記各部は、ソフトウェアおよびハードウェアを併用して実現してもよい。複数のプロセッサを用いる場合、各プロセッサは、各部のうち1つを実現してもよいし、各部のうち2以上を実現してもよい。

[0037] また、上記各部の少なくとも1つは、クラウド上で処理を実行するクラウドサーバに搭載されていてもよい。

[0038] 第1学習部14Aは、話者管理情報12Aから第1変換モデル12Bを学習する。第1変換モデル12Bは、上述したように、得点ベクトルを音響モデルに変換するためのモデルである。

[0039] 第1学習部14Aは、話者管理情報12Aに登録されている複数の話者IDの内、得点ベクトルの登録されている複数の話者IDを特定する。そして、第1学習部14Aは、特定した複数の話者IDの各々に対応する音響モデルと、該話者IDの各々に対応する得点ベクトルと、を用いて、1つの第1

変換モデル 1 2 B を学習する。

[0040] 例えば、第 1 学習部 1 4 A は、非特許文献 2 に開示されている重回帰 H S M M (隠れセミマルコフモデル) や、非特許文献 3 に開示されているクラスタ適応学習などを用いて、話者管理情報 1 2 A から第 1 変換モデル 1 2 B を学習する。

[0041] (非特許文献 2) Tachibana et al. “A technique for controlling voice quality of synthetic speech using multiple regression HSMM,” in Proceedings of INTERSPEECH 2006, pp.2438-2441(2006)

(非特許文献 3) 大谷他、統計的音声合成におけるクラスタ適応学習を利用した知覚表現語による話者制御法の検討、音響学会(春)講演論文集(2016)

[0042] 本実施の形態では、第 1 学習部 1 4 A は、重回帰 H S M M を用いて第 1 学習部 1 4 A を学習する場合を、一例として説明する。しかし、第 1 学習部 1 4 A は、他の手法を用いて、話者管理情報 1 2 A から第 1 変換モデル 1 2 B を学習してもよい。例えば、第 1 学習部 1 4 A は、ベクトルからベクトルへの写像を学習するアルゴリズム、たとえばニューラルネットワークなどを用いて、第 1 変換モデル 1 2 B を学習してもよい。

[0043] 重回帰 H S M M を用いる場合、第 1 変換モデル 1 2 B は、下記式 (1) で表される。

[0044] $\mu = H s + b$. . . 式 (1)

[0045] 式 (1) は、第 1 変換モデル 1 2 B を示す式である。式 (1) 中、 μ は、正規分布で表される音響モデルの平均ベクトルを示す。 μ は、スーパーベクトルと称される場合がある。スーパーベクトルは、正規分布で表される音響モデルの平均ベクトルを、決定木の全リーフノードの平均ベクトルをつなげたベクトルで表したものである。

[0046] また、式 (1) 中、 s は、得点ベクトルを示す。また、 H は、変換行列を示す。 b は、バイアスベクトルを示す。 s (得点ベクトル) は、下記式 (2

) で表される。

[0047] $s = (s_1, s_2, \dots, s_L) \dots$ 式 (2)

[0048] 式 (2) 中、 s は、得点ベクトルを表す。 s_i (i は 1 以上 L 以下の整数) は、得点ベクトルの i 番目の種類の声質の特徴の得点である。 L は、声質の特徴の種類の数である。

[0049] 重回帰 H SMM を用いる場合、第 1 学習部 14 A は、話者管理情報 12 A に含まれる話者 ID の内、得点ベクトルの対応付けられた話者 ID を特定する。そして、第 1 学習部 14 A は、話者管理情報 12 A における、特定した話者 ID に対応する音響モデルの平均ベクトル μ と、該話者 ID に対応する得点ベクトル s と、を学習データとして用いる。第 1 学習部 14 A は、学習データを用いて、変換行列 H と、バイアスベクトル b と、を最尤推定により求める。

[0050] これにより、第 1 学習部 14 A は、式 (1) によって示される第 1 変換モデル 12 B を生成し、記憶部 12 へ登録する。すなわち、第 1 学習部 14 A によって、変換行列 H とバイアスベクトル b の特定された式 (1) が、第 1 変換モデル 12 B として記憶部 12 へ記憶される。

[0051] 第 1 変換モデル 12 B を用いることで、得点ベクトル s から音響モデル (音響モデルの平均ベクトル μ) を導出することができる。すなわち、第 1 変換モデル 12 B を用いることで、処理部 14 は、任意の得点ベクトルによって表される声質の音声を、合成可能となる。

[0052] 次に、第 1 変換部 14 B について説明する。第 1 変換部 14 B は、第 1 変換モデル 12 B を用いて、話者管理情報 12 A に登録されている音響モデルを得点ベクトルに変換し、話者 ID に対応づけて得点管理情報 12 C へ登録する。第 1 変換モデル 12 B は得点ベクトルを音響モデルへ変換するためのモデルである。このため、第 1 変換部 14 B は、第 1 変換モデル 12 B の逆変換モデルを用いて、話者管理情報 12 A に登録されている音響モデルを得点ベクトルに変換し、得点管理情報 12 C へ登録する。

[0053] まず、得点管理情報 12 C について説明する。図 4 は、得点管理情報 12

Cのデータ構成の一例を示す模式図である。

[0054] 得点管理情報12Cは、話者管理情報12Aと同様に、話者IDと、音声データと、音声のテキストと、音響特徴量と、言語特徴量と、音響モデルと、得点ベクトルと、を対応づけたデータベースである。なお、得点管理情報12Cのデータ形式は、データベースに限定されない。例えば、得点管理情報12Cは、テーブルであってもよい。

[0055] 得点管理情報12Cには、話者管理情報12Aに示されるデータの全てが登録されている。そして、得点管理情報12Cには、話者管理情報12Aにおける、得点ベクトルの未登録の話者IDに対応づけて、第1変換部14Bで変換された得点ベクトルが更に登録されている。

[0056] このため、図4に示すように、話者管理情報12A（図2参照）において登録されていた得点ベクトル（S a）は、得点管理情報12Cにそのまま反映される。一方、話者管理情報12Aにおいて、得点ベクトルの未登録の話者ID（例えば、話者ID“C”、“E”）については、得点管理情報12Cでは、第1変換部14Bで変換された得点ベクトル（S b）が登録された状態となる。

[0057] 図1に戻り説明を続ける。本実施の形態では、第1変換部14Bは、話者管理情報12Aにおける、得点ベクトルの未登録の話者IDを特定する。図2に示す話者管理情報12Aの構成の場合、第1変換部14Bは、得点ベクトルの未登録の話者ID（例えば、話者ID“C”、“D”、“E”）を特定する。

[0058] そして、第1変換部14Bは、第1変換モデル12Bの逆変換モデルを用いて、特定した話者IDに対応する音響モデルを、該話者IDに対応する得点ベクトルに変換する。

[0059] 詳細には、第1変換部14Bは、第1変換モデル12Bを示す式（1）を変形し、逆変換モデルを示す下記式（3）を導出する。下記式（3）は、第1変換モデル12Bの逆写像の逆変換モデルを示す式である。

[0060] $s = (H^T H)^{-1} H^T (\mu - b) \quad \cdots \text{式 (3)}$

- [0061] 式(3)中、 s 、 H 、 μ 、および b の定義は、上記式(1)および式(2)と同様である。
- [0062] 第1変換部14Bは、第1変換モデル12Bから、変換行列 H とバイアスベクトル b を特定する。そして、第1変換部14Bは、該変換行列 H とバイアスベクトル b をあてはめた式(3)を、第1変換モデル12Bの逆変換モデルとして用いる。そして、第1変換部14Bは、話者管理情報12Aにおける得点ベクトルの未登録の話者IDに対応する音響モデルの平均ベクトル μ から、得点ベクトル S (図4では得点ベクトル S_b)を導出する。そして、第1変換部14Bは、導出した得点ベクトルを、対応する話者IDに対応づけて、得点管理情報12Cへ登録する。
- [0063] 本実施の形態では、第1変換部14Bは、話者管理情報12Aに登録されているデータの全てを、得点管理情報12Cへ登録した後に、更に、第1変換部14Bで導出した得点ベクトルを、対応する話者IDに対応づけて、得点管理情報12Cへ登録する。
- [0064] このため、上述したように、話者管理情報12A(図2参照)において登録されていた得点ベクトル(S_a)は、得点管理情報12Cにそのまま反映される(図4参照)。一方、話者管理情報12Aにおいて、得点ベクトルの未登録の話者ID(例えば、話者ID“C”、“E”)に対して、得点管理情報12Cでは、第1変換部14Bで変換された得点ベクトル S_b が登録された状態となる。なお、図2および図4に示す例では、話者ID“D”には、音響モデルが対応づけられていないため、第1変換部14Bは、得点ベクトルを導出できない。このため、本実施の形態では、話者ID“D”については、得点管理情報12Cには、得点ベクトルが登録されない。
- [0065] なお、得点管理情報12Cにおける得点ベクトルの全てが、第1変換部14Bで変換された得点ベクトルであってもよい。この場合、第1変換部14Bは、話者管理情報12Aに登録されている全ての話者IDの各々に対応する音響モデルを、第1変換モデル12Bの逆変換モデルを用いて得点ベクトルに変換し、得点管理情報12Cへ登録すればよい。

- [0066] なお、本実施の形態では、話者管理情報 1 2 A と第 1 変換モデル 1 2 B とが別々のデータベースである場合を一例として示した。しかし、話者管理情報 1 2 A と第 1 変換モデル 1 2 B とを、1 つの管理情報（すなわち 1 つのデータベース）として構成してもよい。
- [0067] また、得点管理情報 1 2 C には、少なくとも、話者 I D に対応する得点ベクトルが登録されていればよく、音声データ、音声のテキスト、音響特徴量、言語特徴量、および、音響モデルの少なくとも 1 つが登録されていない形態であってもよい。
- [0068] なお、重回帰 H S M M を用いる場合、第 1 変換部 1 4 B は、上述したように、第 1 変換モデル 1 2 B の逆変換モデルを用いて、音響モデルから得点ベクトルを導出することができる。一方、ニューラルネットワークのように解析的に逆写像が求められない手法を用いる場合、第 1 変換部 1 4 B は、第 1 変換モデル 1 2 B と同様に、音響モデルの平均ベクトル μ から得点ベクトルへの写像をモデル化したモデルを別途学習すればよい。そして、第 1 変換部 1 4 B は、このモデルを用いて、音響モデルから得点ベクトルを導出すればよい。
- [0069] 図 1 に戻り説明を続ける。受付部 1 4 C は、得点ベクトルの入力を受け付ける。ユーザは、入力部 1 8 A を操作することで、検索対象の話者の音声の特徴を示す、得点ベクトルを入力する。上述したように、得点ベクトルは、話者の主観的な特徴を、声質を表す主観的な種類に応じたラベルの付与された軸を用いて表した、ベクトルである。このため、ユーザは、物理的特徴量ではなく、声質を表す主観的なラベルを用いて、検索対象の所望の声質を入力することができる。
- [0070] 得点ベクトルの入力時には、表示制御部 1 4 E の制御によって、表示部 1 8 B には、入力画面が表示される。
- [0071] 表示制御部 1 4 E は、各種画像や情報を表示部 1 8 B へ表示する制御を行う。本実施の形態では、表示制御部 1 4 E は、得点ベクトルの入力を受け付けるための入力画面を表示部 1 8 B へ表示する。

- [0072] 図5は、入力画面30の一例を示す模式図である。入力画面30は、得点ベクトル入力欄30Aと、得点ベクトル表示欄30Bと、再生ボタン30Cと、決定ボタン30Dと、を含む。
- [0073] 得点ベクトル入力欄30Aは、ユーザの所望の、検索対象の話者の声質の特徴を示す得点ベクトルを入力するための入力欄である。例えば、得点ベクトル入力欄30Aには、声質の特徴の種類ごとに、得点を入力するためのスライダバー30Fが表示されている。ユーザは、入力画面30を参照しながら入力部18Aを操作し、スライダバー30Fに示されるつまみ30Eの位置を調整する。この操作によって、ユーザは、声質の特徴の種類ごとに、得点を入力する。
- [0074] 得点ベクトル表示欄30Bは、得点ベクトル入力欄30Aによって入力された得点ベクトルによって示される得点を、声質の特徴の種類ごとに表した表示欄である。
- [0075] 再生ボタン30Cは、得点ベクトル表示欄30Bに表示されている得点ベクトルに応じて合成された音声の再生を指示するときに、ユーザによって操作指示される。決定ボタン30Dは、得点ベクトル表示欄30Bに表示されている得点ベクトルに応じた話者の検索実行を指示するときに、ユーザによって操作指示される。
- [0076] 入力画面30の得点ベクトル入力欄30Aに、ユーザによる入力部18Aの操作指示によって得点ベクトルが入力され、再生ボタン30Cが操作指示される。
- [0077] すると、UI部18は、再生ボタン30Cが操作指示されたときに入力画面30に表示された得点ベクトル表示欄30Bに示される得点の得点ベクトルと、再生指示と、を処理部14へ出力する。
- [0078] これによって、処理部14の受付部14Cは、得点ベクトルと再生指示とを受付ける。
- [0079] 図1に戻り説明を続ける。受付部14Cは、得点ベクトルと再生指示を入力部18Aから受け付けると、受け付けた得点ベクトルを、第2変換部14Fへ

出力する。

[0080] 第2変換部14Fは、受付けた得点ベクトルを、第1変換モデル12Bを用いて音響モデルに変換する。すなわち、第2変換部14Fは、第1変換モデル12Bを用いることで、ユーザによる入力部18Aの操作指示によって入力された得点ベクトルを、該得点ベクトルの表す声質の話者の音響モデルに変換する。そして、第2変換部14Fは、変換した音響モデルを合成部14Gへ出力する。

[0081] 合成部14Gは、第2変換部14Fから受付けた音響モデルから音声を作成する。音響モデルから音声を合成する方法には、公知の方法を用いればよい。

[0082] 例えば、合成部14Gは、音響モデルを用いて、任意のテキストから抽出した言語特徴量を音響特徴量に変換し、ボコーダを用いて、音響特徴量から音声を合成する（例えば、参考文献1参照）。

[0083] （参考文献1）Keiichi Tokuda “Speech Synthesis based on Hidden Markov Models,” in Proceedings of the IEEE, vol. 101, no. 5, pp.1234-1252, 2013.

[0084] ここで、従来手法では、メルケプストラム係数を太さ、F0を声の高さ、周期・非周期成分をかすれ、など、物理特徴量と得点を一対一に対応させることがよく行われていた。一方、本実施の形態では、第1学習部14Aが、専用の変換モデルである第1変換モデル12Bを、話者管理情報12Aから学習する。このため、合成部14Gは、第1変換モデル12Bを用いることで、任意の話者の音響モデルから、人が主観的に感じる特徴の声質の音声を合成できる。

[0085] 再生部14Hは、合成部14Gで合成された音声を再生するように音声出力部16を制御する。音声出力部16は、合成部14Gで合成された音声を再生する。このため、ユーザは、入力部18Aの操作指示によって入力した得点ベクトルに応じた音声を確認することができる。また、ユーザは、入力画面30の操作時に、再生ボタン30Cを操作することで、入力中の得点ベ

クトルに応じた音声を確認しながら、検索対象の所望の得点ベクトルを入力することができる。

[0086] そして、ユーザは、検索対象として用いる得点ベクトルが定まったときに、入力画面30における決定ボタン30Dを操作指示すればよい。

[0087] すなわち、入力画面30の得点ベクトル入力欄30Aに、ユーザによる入力部18Aの操作指示によって得点ベクトルが入力され、決定ボタン30Dが操作指示される。すると、UI部18は、決定ボタン30Dが操作指示されたときに入力画面30に表示された得点ベクトル表示欄30Bに示される得点の得点ベクトルと、検索実行指示と、を処理部14へ出力する。

[0088] 処理部14の受付部14Cは、得点ベクトルと検索実行指示とを受付ける。

[0089] 受付部14Cは、得点ベクトルと検索実行指示とをUI部18から受け付けると、該得点ベクトルを検索部14Dへ出力する。

[0090] 検索部14Dは、得点管理情報12Cから、受け付けた得点ベクトルに類似する得点ベクトルに対応する話者IDを、検索する。

[0091] 詳細には、検索部14Dは、ユークリッド距離などを用いて、入力部18Aから受付部14Cを介して受け付けた得点ベクトルと、得点管理情報12Cに登録されている得点ベクトルの各々と、の距離を計算する。ユークリッド距離は、例えば、下記式(4)で表される。

[0092] [数1]

$$d(s, t) = \sqrt{\sum_{i=1}^L (s_i - t_i)^2} \quad \dots \text{式(4)}$$

[0093] 式(4)中、sは、受付部14Cを介して入力部18Aから受け付けた、検索対象(すなわち、検索クエリ(キー))の得点ベクトルを示す。sは、上記式(2)で表される。また、式(4)中、s_i(iは1以上L以下の整数)は、受け付けた得点ベクトルのi番目の種類の声質の特徴の得点である。

[0094] また、式(4)中、t_i(iは1以上L以下の整数)は、得点管理情報12Cに登録されている得点ベクトルのi番目の種類の声質の特徴の得点であ

る。なお、得点管理情報 1 2 C に登録されている得点ベクトルと、受付部 1 4 C で入力部 1 8 A から受付けた得点ベクトルと、の声質の特徴の種類は同じであるものとする。

[0095] また、式 (4) 中、 t は、得点管理情報 1 2 C に登録されている得点ベクトルを示す。 t は、下記式 (5) で表される。

[0096] $t = (t_1, t_2, \dots, t_L) \dots$ 式 (5)

[0097] 式 (5) 中、 L は、上記と同様に、声質の特徴の種類の数である。

[0098] そして、検索部 1 4 D は、得点管理情報 1 2 C に登録されている得点ベクトルの内、入力部 1 8 A から受付部 1 4 C を介して受付けた得点ベクトルとの距離の近い上位 K 個の得点ベクトルを特定する (K は 1 以上の整数)。そして、検索部 1 4 D は、得点管理情報 1 2 C における、特定した該得点ベクトルに対応する話者 $I D$ を特定する。これによって、検索部 1 4 D は、得点管理情報 1 2 C から、受付部 1 4 C で受付けた得点ベクトルに類似する話者 $I D$ を検索する。

[0099] すなわち、本実施の形態では、検索部 1 4 D は、メルケプストラム係数や $F0$ などの物理特徴量で話者の声間の類似度を求めるのではなく、これらの人の主観的な判断を反映した得点ベクトルを用いて、類似する話者 $I D$ を検索する。

[0100] なお、検索部 1 4 D は、声質の特徴の種類に対する重み付けを行った得点ベクトルを用いて、話者 $I D$ を検索してもよい。

[0101] この場合、表示制御部 1 4 E は、入力画面 3 0 に、声質の特徴の種類ごとに、得点を入力するための入力領域 (図 5 ではスライダーバー 3 0 F) と共に、重み付け値の入力欄を設け、表示部 1 8 B へ表示すればよい。

[0102] そして、ユーザは、入力画面 3 0 を参照しながら入力部 1 8 A を操作することで、声質の特徴の種類ごとに、重み付け値を更に入力すればよい。この場合、受付部 1 4 C は、声質の特徴の種類ごとの得点を示す得点ベクトルと、声質の特徴の種類に対する重み付け値と、検索実行指示と、を受付ける。

[0103] そして、検索部 1 4 D は、受付けた前記得点ベクトルに含まれる、声質の

特徴の種類ごとの得点を、対応する重み付け値で重み付けした得点ベクトルを、検索クエリ（キー）として用いて、上記と同様にして、得点管理情報 12C から類似する話者 ID を検索すればよい。

[0104] 詳細には、検索部 14D は、下記式（6）に示されるユークリッド距離を用いて、入力部 18A から受付部 14C を介して受付けた得点ベクトルと、得点管理情報 12C に登録されている得点ベクトルの各々と、の距離を計算する。

[0105] [数2]

$$dw(s, t) = \sqrt{\sum_{i=1}^L w_i (s_i - t_i)^2} \quad \cdots \text{式(6)}$$

[0106] 式（6）中、 s 、 s_i 、 t 、 t_i は、上記式（4）と同様である。式（6）中、 w は、重み付け値を示す。式（6）中、 w_i は、 i 番目の種類の声質の特徴に対する、重み付け値を示す。

[0107] このように、検索部 14D は、声質の特徴の種類に対する重み付けを行った得点ベクトルを用いて、話者 ID を検索してもよい。

[0108] 重み付けを行うことで、ユーザの所望の話者 ID が検索されやすくなる。なお、処理部 14 は、人の主観的な声の類似尺度に合うように、声質の特徴の種類に対する最適化した重み付けを、予め行っておいてもよい。

[0109] 検索部 14D は、検索した話者 ID、および、得点管理情報 12C における該話者 ID に対応する得点ベクトルを、表示制御部 14E へ出力する。表示制御部 14E は、検索部 14D で検索された話者 ID を含む表示画面を、表示部 18B へ表示する。

[0110] 図 6 は、表示画面 32 の一例を示す模式図である。表示画面 32 は、得点ベクトル表示欄 32A と、得点ベクトル表示欄 32B と、検索結果表示欄 32C と、再生ボタン 32D と、を含む。

[0111] 得点ベクトル表示欄 32A および得点ベクトル表示欄 32B は、検索クエリとして用いた得点ベクトルの表示欄である。検索結果表示欄 32C は、得点ベクトル表示欄 32A および得点ベクトル表示欄 32B に表示された得点

ベクトルを用いて検索された、話者 I D の検索結果の表示欄である。

[0112] 図 6 に示す例では、検索結果表示欄 3 2 C には、検索部 1 4 D で検索された話者 I D の一覧と、各話者 I D に対応する得点ベクトルと、が表示される。なお、図 6 には、上記 K（受付けた得点ベクトルとの距離の近い上位 K 個）が、“3”である場合を一例として示している。すなわち、図 6 には、得点ベクトル表示欄 3 2 A および得点ベクトル表示欄 3 2 B に示される得点ベクトルに類似する、得点ベクトルに対応する K 個の話者 I D として、3 個の話者 I D を検索した例を示している。

[0113] 検索結果表示欄 3 2 C における、各話者 I D の各々に対応する領域には、再生ボタン 3 2 D（3 2 D 1 ～ 3 2 D 3）が表示されている。再生ボタン 3 2 D は、対応する話者 I D の得点ベクトルに応じて合成された音声の再生を指示するときに、ユーザによって操作指示される。

[0114] ユーザは、入力部 1 8 A を操作することで、再生ボタン 3 2 D を操作指示する。すると、UI 部 1 8 は、該再生ボタン 3 2 D に対応する得点ベクトルと、再生指示と、を処理部 1 4 へ出力する。例えば、再生ボタン 3 2 D 1 が操作指示されたと仮定する。すると、UI 部 1 8 は、該再生ボタン 3 2 D 1 に対応する、話者 I D “A” に対応する得点ベクトル（検索結果表示欄 3 2 C に表示されている得点ベクトル）と、再生指示と、を処理部 1 4 へ出力する。

[0115] 処理部 1 4 は、UI 部 1 8 から、得点ベクトルと再生指示を受付けると、受付部 1 4 C、第 2 変換部 1 4 F、合成部 1 4 G、および再生部 1 4 H が、上記と同様の処理を行う。このため、ユーザによって操作指示された再生ボタン 3 2 D に対応する話者 I D の得点ベクトルが、音響モデルに変換され、該音響モデルから合成された音声、音声出力部 1 6 から再生される。

[0116] このため、ユーザは、検索結果表示欄 3 2 C に示される各話者 I D の各々に対応する音声を確認することができる。

[0117] なお、検索部 1 4 D は、検索結果を、ネットワークなどの通信回線を介して外部装置へ送信してもよい。この場合、検索部 1 4 D は、検索クエリとし

て用いた得点ベクトルと、検索した話者の話者IDと、該話者IDに対応する得点ベクトルと、を外部装置へ送信する。これにより、これらの情報を受付けた外部装置側で、入力した得点ベクトルに応じた話者の検索結果を確認することが可能となる。

[0118] 次に、話者検索装置10で実行する検索処理の手順の一例を説明する。

[0119] 図7は、検索処理の手順の一例を示す、フローチャートである。まず、第1学習部14Aが、話者管理情報12Aを用いて第1変換モデル12Bを学習する（ステップS100）。第1学習部14Aは、学習した第1変換モデル12Bを、記憶部12へ記憶する（ステップS102）。

[0120] 次に、第1変換部14Bが、話者管理情報12Aに登録されている音響モデルを、第1変換モデル12Bを用いて得点ベクトルに変換し（ステップS104）、得点管理情報12Cへ登録する（ステップS106）。次に、処理部14は、話者検索処理を実行する（ステップS108）。そして、本ルーチンを終了する。

[0121] 図8は、話者検索処理（図7のステップS108）の手順の一例を示すフローチャートである。

[0122] まず、処理部14は、記憶部12に記憶されている第1変換モデル12Bを読み込む（ステップS200）。次に、表示制御部14Eが、入力画面30をUI部18に表示する（ステップS202）。ステップS202の処理によって、UI部18には、例えば、図5に示す入力画面30が表示される。

[0123] ユーザは、入力画面30を参照しながら入力部18Aを操作し、声質の特徴の種類ごとに得点を入力する。これによって、ユーザは、検索対象の話者の声質の特徴に応じた得点ベクトルを入力する。そして、ユーザは、再生ボタン30Cを操作指示する。

[0124] 次に、受付部14Cは、得点ベクトルの入力を受付けたと判断するまで（ステップS204：Yes）、否定判断を繰返す（ステップS204：No）。例えば、受付部14Cは、入力部18Aから得点ベクトルおよび再生指示を受付けたか否かを判別することで、ステップS204の判断を行う。

- [0125] ステップS204で肯定判断すると（ステップS204：Yes）、ステップS206へ進む。ステップS206では、第2変換部14Fが、ステップS204で受付けた得点ベクトルを、第1変換モデル12Bを用いて音響モデルに変換する（ステップS206）。次に、合成部14Gが、ステップS206で変換された音響モデルから音声を合成する（ステップS208）。そして、再生部14Hが、ステップS208で合成された音声を入力するように、音声出力部16を再生制御する（ステップS210）。
- [0126] ステップS210の処理によって、ユーザの操作によって入力された得点ベクトルに応じた音声、音声出力部16から出力される。
- [0127] 次に、受付部14Cは、検索実行指示を受付けたか否かを判断する（ステップS212）。ステップS212では、受付部14Cは、得点ベクトルおよび検索実行指示を、入力部18Aから受付けたか否かを判別することで、ステップS212の判断を行う。なお、ステップS204で否定判断した場合、ステップS212の判断を行ってもよい。
- [0128] ステップS212で否定判断すると（ステップS212：No）、上記ステップS204へ戻る。一方、ステップS212で肯定判断すると（ステップS212：Yes）、ステップS214へ進む。
- [0129] ステップS214では、検索部14Dが、得点管理情報12Cから、ステップS212で受付けた得点ベクトルに類似する得点ベクトルに対応する話者IDを検索する（ステップS214）。
- [0130] 次に、表示制御部14Eが、ステップS214で検索された話者IDを含む表示画面32を、表示部18Bへ表示する（ステップS216）。ステップS216の処理によって、例えば、表示部18Bには、図6に示す表示画面32が表示される。
- [0131] 次に、受付部14Cが、入力部18Aから再生指示を受付けたか否かを判断する（ステップS218）。ステップS218で否定判断すると（ステップS218：No）、後述するステップS226へ進む。一方、ステップS218で肯定判断すると（ステップS218：Yes）、ステップS220

へ進む。

- [0132] ステップS 2 2 0では、第2変換部1 4 Fが、ステップS 2 1 8で再生指示と共に受付けた得点ベクトルを、第1変換モデル1 2 Bを用いて音響モデルに変換する（ステップS 2 2 0）。次に、合成部1 4 Gが、ステップS 2 2 0で変換された音響モデルから音声を合成する（ステップS 2 2 2）。そして、再生部1 4 Hが、ステップS 2 2 2で合成された音声を出力するように、音声出力部1 6を再生制御する（ステップS 2 2 4）。
- [0133] ステップS 2 2 4の処理によって、ユーザの操作によって入力された得点ベクトルに応じた音声が、音声出力部1 6から出力される。
- [0134] 次に、処理部1 4は、処理を終了するか否かを判断する（ステップS 2 2 6）。処理部1 4は、例えば、ユーザによる入力部1 8 Aの操作指示によって、処理終了を指示するための予め定められたボタンが操作されたか否かを判別することで、ステップS 2 2 6の判断を行う。ステップS 2 2 6で否定判断すると（ステップS 2 2 6：N o）、上記ステップS 2 1 8へ戻る。なお、ステップS 2 2 6で否定判断した場合、上記ステップS 2 0 2へ戻ってもよい。ステップS 2 2 6で肯定判断すると（ステップS 2 2 6：Y e s）、本ルーチンを終了する。
- [0135] 以上説明したように、本実施の形態の話者検索装置1 0は、第1変換部1 4 Bと、受付部1 4 Cと、検索部1 4 Dと、を備える。第1変換部1 4 Bは、声質の特徴を表す得点ベクトルを音響モデルに変換するための第1変換モデル1 2 Bの逆変換モデルを用いて、予め登録された音響モデルを得点ベクトルに変換し、話者I D（話者識別情報）に対応づけて得点管理情報1 2 Cに登録する。受付部1 4 Cは、得点ベクトルの入力を受付ける。検索部1 4 Dは、得点管理情報1 2 Cから、受付けた得点ベクトルに類似する話者I D（話者識別情報）を検索する。
- [0136] ここで、従来では、ユーザ所望の話者の音声データを、検索クエリとして用いて、ユーザ所望の音声の話者を検索していた。このため、従来では、ユーザ所望の話者の音声データを事前に登録しておかなければ、話者を検索す

ることができなかった。

[0137] 一方、本実施の形態の話者検索装置 10 は、声質の特徴を表す得点ベクトルを検索クエリとして用いて、得点管理情報 12 C から、受付けた得点ベクトルに類似する話者 ID を検索する。また、本実施の形態の話者検索装置 10 は、声質の特徴を表す得点ベクトルを音響モデルに変換するための第 1 変換モデル 12 B を用いて、予め登録された音響モデルを得点ベクトルに変換し、話者 ID（話者識別情報）に対応づけて得点管理情報 12 C に登録する。

[0138] 従って、本実施の形態の話者検索装置 10 は、声質類似度の高い話者を検索することができる。

[0139] また、従来では、膨大な話者の各々に対応する音声データを、予めカテゴリごとに分類し、所望の話者を検索する手法がとられていた。しかし、登録する話者の数が膨大になるほど、分類作業の負荷やコストも膨大となっていた。

[0140] 一方、本実施の形態の話者検索装置 10 は、声質の特徴を表す得点ベクトルを検索クエリとして用いて、得点管理情報 12 C から、受付けた得点ベクトルに類似する話者 ID を検索する。

[0141] このため、本実施の形態の話者検索装置 10 は、検索の母体となるデータベースに登録されている話者の音声を予め分類する負荷やコストの増大を招くことなく、声質類似度の高い話者を検索することができる。

[0142] また、従来では、話者間の類似度を、メル周波数ケプストラム係数、基本周波数（F0）などの基本的な物理特徴量間の比較によって導出していた。このため、従来では、これらの物理特徴量間の類似度と人が主観的に感じる話者間の近さが一致しない場合があった。

[0143] 一方、本実施の形態の話者検索装置 10 において、検索クエリとして用いる得点ベクトルは、話者の声質の主観的な特徴を表す。このため、本実施の形態の話者検索装置 10 は、上記効果に加えて、主観的な声質類似度の高い話者を検索することができる。

[0144] (第2の実施の形態)

上記実施の形態では、得点管理情報12Cには、話者管理情報12Aに登録されている全てのデータと、第1変換部14Bで登録された得点データと、が登録されている場合を示した。

[0145] すなわち、上記実施の形態では、得点管理情報12Cには、予め話者管理情報12Aに登録されていた得点データと、第1変換部14Bによって音響モデルから導出された得点データと、が登録されている場合を示した。

[0146] 本実施の形態では、得点管理情報12Cに、音響特徴量から導出した得点ベクトルを更に登録する形態を説明する。

[0147] 図9は、本実施の形態の話者検索装置10Aの一例を示す図である。

[0148] 話者検索装置10Aは、記憶部13と、処理部15と、音声出力部16と、UI部18と、を備える。記憶部13、音声出力部16、およびUI部18と、処理部15と、は、バスなどを介してデータや信号を授受可能に接続されている。音声出力部16およびUI部18は、第1の実施の形態と同様である。

[0149] 記憶部13は、各種データを記憶する。記憶部13は、例えば、RAM、フラッシュメモリ等の半導体メモリ素子、ハードディスク、光ディスク等である。なお、記憶部13は、話者検索装置10Aの外部に設けられた記憶装置であってもよい。また、記憶部13は、記憶媒体であってもよい。また、記憶部13を、複数の記憶媒体から構成してもよい。

[0150] 本実施の形態では、記憶部13は、話者管理情報12Aと、第1変換モデル12Bと、得点管理情報13Cと、第3変換モデル13Dと、音声管理情報13Eと、を記憶する。話者管理情報12Aおよび第1変換モデル12Bは、第1の実施の形態と同様である。

[0151] 音声管理情報13Eは、話者の音声に関する情報である。音声管理情報13Eは、第3変換モデル13Dの学習に用いるための音声に関する情報が登録されている。音声管理情報13Eは、記憶部13に予め記憶されている。

[0152] 図10は、音声管理情報13Eのデータ構成の一例を示す模式図である。

音声管理情報 1 3 E は、話者 I D と、音声データと、を対応づけたデータベースである。なお、音声管理情報 1 3 E のデータ形式は、データベースに限定されない。例えば、音声管理情報 1 3 E は、テーブルであってもよい。

[0153] 話者 I D は、第 1 の実施の形態と同様に、話者を識別する識別情報（話者識別情報）である。なお、音声管理情報 1 3 E には、話者管理情報 1 2 A（図 2 参照）に登録されている話者 I D の内、少なくとも音響モデルの対応付けられていない話者 I D（例えば、話者 I D “D”）が登録されている。また、音声管理情報 1 3 E には、話者管理情報 1 2 A に登録されていない話者の話者 I D（例えば、話者 I D “X”、“Y”、“Z”）が登録されている。

[0154] なお、音声管理情報 1 3 E には、上記話者 I D に加えて、更に、話者管理情報 1 2 A に登録されている話者 I D の少なくとも一部も登録されていてもよい。

[0155] 音声管理情報 1 3 E には、話者 I D に対応づけて、複数の音声データが登録されている。なお、音声管理情報 1 3 E における、各話者 I D に対応して登録されている音声データの数は、話者管理情報 1 2 A の各話者 I D に対応して登録されている音声データの数（N 個）より少ない数であって、且つ、1 より多い数であればよい。

[0156] 図 9 に戻り説明を続ける。第 3 学習部 1 5 J は、話者管理情報 1 2 A から第 3 変換モデル 1 3 D を学習する。第 3 変換モデル 1 3 D は、音響特徴量を得点ベクトルに変換するためのモデルである。

[0157] 第 3 学習部 1 5 J は、話者管理情報 1 2 A に登録されている話者 I D の各々に対応する音響特徴量と、該話者管理情報 1 2 A における話者 I D の各々に対応する得点ベクトルと、を用いて、第 3 変換モデル 1 3 D を学習する。

[0158] 第 3 学習部 1 5 J は、音響特徴量のベクトルから得点ベクトルへの写像を学習する公知の手法を用いて、第 3 変換モデル 1 3 D を学習すればよい。例えば、第 3 学習部 1 5 J は、線形回帰モデル、カーネル回帰モデル、ニューラルネットワークなどを用いて、第 3 変換モデル 1 3 D を学習する。なお、

第3学習部15Jは、これらのモデルのパラメータを、話者管理情報12Aに含まれる話者IDの各々に対応する音響特徴量および得点ベクトルを用いて、推定すればよい。

[0159] そして、第3学習部15Jは、学習した第3変換モデル13Dを、記憶部13へ記憶する。

[0160] 特徴抽出部15Kは、音声管理情報13Eに登録されている音声データから、音響特徴量を抽出する。詳細には、特徴抽出部15Kは、音声管理情報13Eに登録されている話者IDごとに、音声データから音響特徴量を抽出する。音響特徴量は、ベクトルで表される。なお、特徴抽出部15Kは、公知の方法を用いて、音声データから音響特徴量を抽出すればよい。

[0161] なお、音響特徴量は、音響モデルの生成に比べて、少ない数の音声データから抽出することできる。このため、上述したように、音声管理情報13Eにおける、各話者IDに対応して登録されている音声データの数、話者管理情報12Aの各話者IDに対応して登録されている音声データの数（N個）より少ない数であって、且つ、1より多い数であればよい。

[0162] 第3変換部15Lは、第3変換モデル13Dを用いて、話者の音響特徴量を得点ベクトルに変換し、得点管理情報13Cへ登録する。

[0163] 詳細には、第3変換部15Lは、音声管理情報13Eから特徴抽出部15Kによって、話者IDごとに抽出された音声特徴量を、第3変換モデル13Dを用いて得点ベクトルに変換する。そして、第3変換部15Lは、音声管理情報13Eに登録されている話者IDと、変換した得点ベクトルと、を対応付けて、得点管理情報13Cへ登録する。

[0164] 得点管理情報13Cは、得点管理情報12Cと同様に、ユーザによる話者の検索時に用いられるデータベースである。得点管理情報13Cは、第1変換部14Bおよび第3変換部15Lによって登録および更新される。

[0165] 図11は、得点管理情報13Cのデータ構成の一例を示す模式図である。得点管理情報13Cは、話者管理情報12Aおよび第1の実施の形態の得点管理情報12Cと同様に、話者IDと、音声データと、音声のテキストと、

音響特徴量と、言語特徴量と、音響モデルと、得点ベクトルと、を対応づけたデータベースである。なお、得点管理情報 1 3 C のデータ形式は、データベースに限定されない。例えば、得点管理情報 1 3 C は、テーブルであってもよい。

[0166] 得点管理情報 1 3 C は、話者管理情報 1 2 A における、得点ベクトルの未登録の話者 I D に対応づけて、第 1 変換部 1 4 B で変換された得点ベクトルと、第 3 変換部 1 5 L で変換された得点ベクトルと、を更に登録したものである。

[0167] ここで、第 1 の実施の形態で用いた得点管理情報 1 2 C (図 4 参照) には、予め話者管理情報 1 2 A に登録されていた得点ベクトル S a と、第 1 変換部 1 4 B によって音響モデルから変換された得点ベクトル S b と、が登録されている。このため、得点管理情報 1 2 C には、話者管理情報 1 2 A において音響モデルの未登録の話者 I D (例えば、話者 I D “D”) については、得点ベクトルが登録されていなかった (図 4 参照)。

[0168] 一方、本実施の形態で用いる得点管理情報 1 3 C には、図 1 1 に示すように、第 1 の実施の形態における得点管理情報 1 2 C (図 4 参照) に登録されている得点ベクトル (得点ベクトル (S a)、(S b)) に加えて、音響特徴量から変換された得点ベクトル (S c) が登録されている。このため、得点管理情報 1 3 C には、話者管理情報 1 2 A において音響モデルが未登録であった話者 I D (例えば、話者 I D “D”) に対する、得点ベクトルも、登録される。

[0169] また、音声管理情報 1 3 E には、話者管理情報 1 2 A に未登録の話者 I D に対応する音声データが登録されている。そして、第 3 変換部 1 5 L は、この音声管理情報 1 3 E から特徴抽出部 1 5 K によって抽出された音響特徴量を用いて、得点ベクトルを導出し、得点管理情報 1 3 C へ登録する。

[0170] このため、得点管理情報 1 3 C には、第 1 の実施の形態で用いた得点管理情報 1 2 C に比べて、より多い話者の話者 I D に対応する得点ベクトルが登録され、類似話者検索の対象とすることができる。

[0171] また、第１の実施の形態における、第１変換部１４Ｂによって登録された得点ベクトル（Ｓｂ）は、音響モデルから導出されたものである。このため、音響モデルから導出された得点ベクトルを用いて、第２変換部１４Ｆおよび合成部１４Ｇによって合成される音声は、合成音（実際の話者の肉声ではない合成された音声）である。

[0172] 一方、第３変換部１５Ｌによって登録される得点ベクトルは、音声管理情報１３Ｅから特徴抽出部１５Ｋによって抽出された音響特徴量から、第３変換モデル１３Ｄを用いて導出されたものである。このため、音響特徴量から導出された得点ベクトルを用いて、第２変換部１４Ｆおよび合成部１４Ｇによって合成される音声は、実際の話者の肉声または肉声により近い音声となる。

[0173] このため、検索部１４Ｄが、得点管理情報１２Ｃに代えて得点管理情報１３Ｃから、受付部１４Ｃで受付けた得点ベクトルに類似する、得点ベクトルに対応する話者ＩＤを検索することで、検索部１４Ｄは、合成音に加えて肉声の音声の話者を検索することができる。

[0174] また、本実施の形態の話者検索装置１０Ａでは、共通の得点ベクトルを用いて、合成音および肉声の双方の音声の話者を検索することができる。このため、合成音と肉声を横断的に検索することができる。

[0175] このため、例えば、ユーザは、所望の合成音の音質に満足できれば、音響モデルを使用する。そして、音質に不満であれば、ユーザは、入力した得点ベクトルに類似する肉声の音質の話者ＩＤの検索結果を用いて、該話者ＩＤによって識別される話者に収録を依頼する、といった使用方法も可能である。

[0176] 次に、話者検索装置１０Ａで実行する検索処理の手順の一例を説明する。図１２は、話者検索装置１０Ａが実行する検索処理の手順の一例を示す、フローチャートである。

[0177] まず、処理部１５は、ステップＳ３００～ステップＳ３０６の処理を実行する。ステップＳ３００～ステップＳ３０６の処理は、第１の実施の形態の

ステップS 1 0 0～ステップS 1 0 6と同様である（図7参照）。

[0178] 具体的には、まず、第1学習部14Aが、話者管理情報12Aを用いて第1変換モデル12Bを学習する（ステップS 3 0 0）。第1学習部14Aは、学習した第1変換モデル12Bを、記憶部13へ記憶する（ステップS 3 0 2）。

[0179] 次に、第1変換部14Bが、話者管理情報12Aに登録されている音響モデルを、第1変換モデル12Bを用いて得点ベクトルに変換し、得点管理情報13Cへ登録する（ステップS 3 0 6）。

[0180] 次に、第3学習部15Jが、話者管理情報12Aを用いて第3変換モデル13Dを学習する（ステップS 3 0 8）。そして、第3学習部15Jは、学習した第3変換モデル13Dを記憶部13へ記憶する（ステップS 3 1 0）。

[0181] 次に、特徴抽出部15Kが、音声管理情報13Eに登録されている音声データの各々に対応する音響特徴量を抽出する（ステップS 3 1 2）。次に、第3変換部15Lが、ステップS 3 1 2で話者IDごとに抽出された音響特徴量の各々を、第3変換モデル13Dを用いて得点ベクトルに変換する（ステップS 3 1 4）。

[0182] 次に、第3学習部15Jは、ステップS 3 1 4で変換した得点ベクトルを、該得点ベクトルに対応する話者IDに対応づけて、得点管理情報13Cへ登録する（ステップS 3 1 6）。

[0183] 次に、処理部15は、話者検索処理を実行する（ステップS 3 1 6）。ステップS 3 1 6の処理は、第1の実施の形態（図7のステップS 1 0 8、および図8参照）と同様である。そして、本ルーチンを終了する。

[0184] 以上説明したように、本実施の形態の話者検索装置10Aは、第1の実施の形態の構成に加えて、第3変換部15Lを備える。第3変換部15Lは、音響特徴量を得点ベクトルに変換するための第3変換モデル13Dを用いて、話者の音響特徴量を得点ベクトルに変換し、得点管理情報13Cへ登録する。

- [0185] このため、本実施の形態の話者検索装置 10A では、合成音に加えて肉声の音声の話者についても、検索対象とすることができる。
- [0186] 従って、本実施の形態の話者検索装置 10A は、第 1 の実施の形態の効果に加えて、更に、声質類似度の高い話者を検索することができる。
- [0187] また、本実施の形態の話者検索装置 10A では、第 3 変換部 15L が、メルケプストラムや基本周波数などの言語に存しない音響特徴量のみを用いて、第 3 変換モデル 13D を学習する。このため、話者検索装置 10A では、言語に依存せずに、類似する話者を検索することが可能になる。
- [0188] 次に、本実施の形態の話者検索装置 10、10A のハードウェア構成を説明する。図 13 は、本実施の形態の話者検索装置 10、10A のハードウェア構成例を示す説明図である。
- [0189] 本実施の形態の話者検索装置 10、10A は、CPU 86 などの制御装置と、ROM (Read Only Memory) 88 や RAM (Random Access Memory) 90 などの記憶装置と、ネットワークに接続して通信を行う通信 I/F 82 と、音声を出力するスピーカ 80 と、表示部 98 と、入力部 94 と、HDD 92 と、各部を接続するバス 96 を備えている。
- [0190] 本実施の形態の話者検索装置 10、10A で実行されるプログラムは、ROM 88 等に予め組み込まれて提供される。
- [0191] 本実施の形態の話者検索装置 10、10A で実行されるプログラムは、インストール可能な形式又は実行可能な形式のファイルで CD-ROM (Compact Disk Read Only Memory)、フレキシブルディスク (FD)、CD-R (Compact Disk Recordable)、DVD (Digital Versatile Disk) 等のコンピュータで読み取り可能な記録媒体に記録してコンピュータプログラムプロダクトとして提供されるように構成してもよい。
- [0192] さらに、本実施の形態の話者検索装置 10、10A で実行されるプログラムを、インターネット等のネットワークに接続されたコンピュータ上に格納

し、ネットワーク経由でダウンロードさせることにより提供するように構成してもよい。また、本実施の形態にかかる話者検索装置 10、10A で実行されるプログラムをインターネット等のネットワーク経由で提供または配布するように構成してもよい。

[0193] 本実施の形態の話者検索装置 10、10A で実行されるプログラムは、コンピュータを上述した話者検索装置 10、10A の各部として機能させうる。このコンピュータは、CPU 86 がコンピュータで読取可能な記憶媒体からプログラムを主記憶装置上に読み出して実行することができる。

[0194] なお、上記実施の形態では、話者検索装置 10、10A の各々が、単体の装置として構成されていることを想定して説明した。しかし、話者検索装置 10、10A の各々は、物理的に分離されてネットワークを介して接続された複数の装置により構成されていてもよい。また、上記実施の形態の話者検索装置 10、10A は、クラウドシステム上で動作する仮想マシンとして実現されていてもよい。

[0195] なお、上記には、本発明の実施の形態を説明したが、上記実施の形態は、例として提示したものであり、発明の範囲を限定することは意図していない。この新規な実施の形態は、その他の様々な形態で実施されることが可能であり、発明の要旨を逸脱しない範囲で、種々の省略、置き換え、変更を行うことができる。この実施の形態やその変形は、発明の範囲や要旨に含まれるとともに、請求の範囲に記載された発明とその均等の範囲に含まれる。

符号の説明

- [0196] 10、10A 話者検索装置
12、13 記憶部
12A 話者管理情報
12B 第1変換モデル
12C、13C 得点管理情報
13D 第3変換モデル
13E 音声管理情報

1 4、1 5 処理部

1 4 A 第1 学習部

1 4 B 第1 変換部

1 4 C 受付部

1 4 D 検索部

1 4 E 表示制御部

1 4 F 第2 変換部

1 4 G 合成部

1 4 H 再生部

1 5 J 第3 学習部

1 5 K 特徴抽出部

1 5 L 第3 変換部

1 6 音声出力部

1 8 U I 部

1 8 A 入力部

1 8 B 表示部

請求の範囲

- [請求項1] 声質の特徴を表す得点ベクトルを音響モデルに変換するための第1変換モデルの逆変換モデルを用いて、予め登録された前記音響モデルを前記得点ベクトルに変換し、話者識別情報に対応づけて得点管理情報に登録する第1変換部と、
前記得点ベクトルの入力を受付ける受付部と、
前記得点管理情報から、受付けた前記得点ベクトルに類似する話者識別情報を検索する検索部と、
を備える話者検索装置。
- [請求項2] 前記得点ベクトルは、
話者の声質の主観的な特徴を表す、
請求項1に記載の話者検索装置。
- [請求項3] 受付けた前記得点ベクトルを、前記第1変換モデルを用いて前記音響モデルに変換する第2変換部と、
前記音響モデルから合成された音声を再生する再生部と、
を備える、請求項1に記載の話者検索装置。
- [請求項4] 前記音響モデルは、音声の言語特徴量から音声の音響特徴量への写像を統計的にモデル化する、請求項1に記載の話者検索装置。
- [請求項5] 前記音響特徴量を前記得点ベクトルに変換するための第3変換モデルを用いて、話者の前記音響特徴量を前記得点ベクトルに変換し、前記得点管理情報に登録する第3変換部、
を備える、請求項4に記載の話者検索装置。
- [請求項6] 前記音響特徴量を前記得点ベクトルに変換するための前記第3変換モデルを学習する第3学習部を備える、
請求項5に記載の話者検索装置。
- [請求項7] 前記得点ベクトルは、
声質の特徴の主観的な強さを、声質の特徴の種類ごとに得点で表し、

前記受付部は、

前記特徴の種類ごとの得点を示す前記得点ベクトルと、前記特徴の種類に対する重み付け値と、を受付け、

前記検索部は、

受付けた前記得点ベクトルに含まれる前記得点を対応する重み付け値で重み付けした前記得点ベクトルに類似する、話者識別情報を検索する、

請求項 1 に記載の話者検索装置。

[請求項8]

声質の特徴を表す得点ベクトルを音響モデルに変換するための第 1 変換モデルの逆変換モデルを用いて、予め登録された前記音響モデルを前記得点ベクトルに変換し、話者識別情報に対応づけて得点管理情報に登録するステップと、

前記得点ベクトルの入力を受付けるステップと、

前記得点管理情報から、受付けた前記得点ベクトルに類似する話者識別情報を検索するステップと、

を含む話者検索方法。

[請求項9]

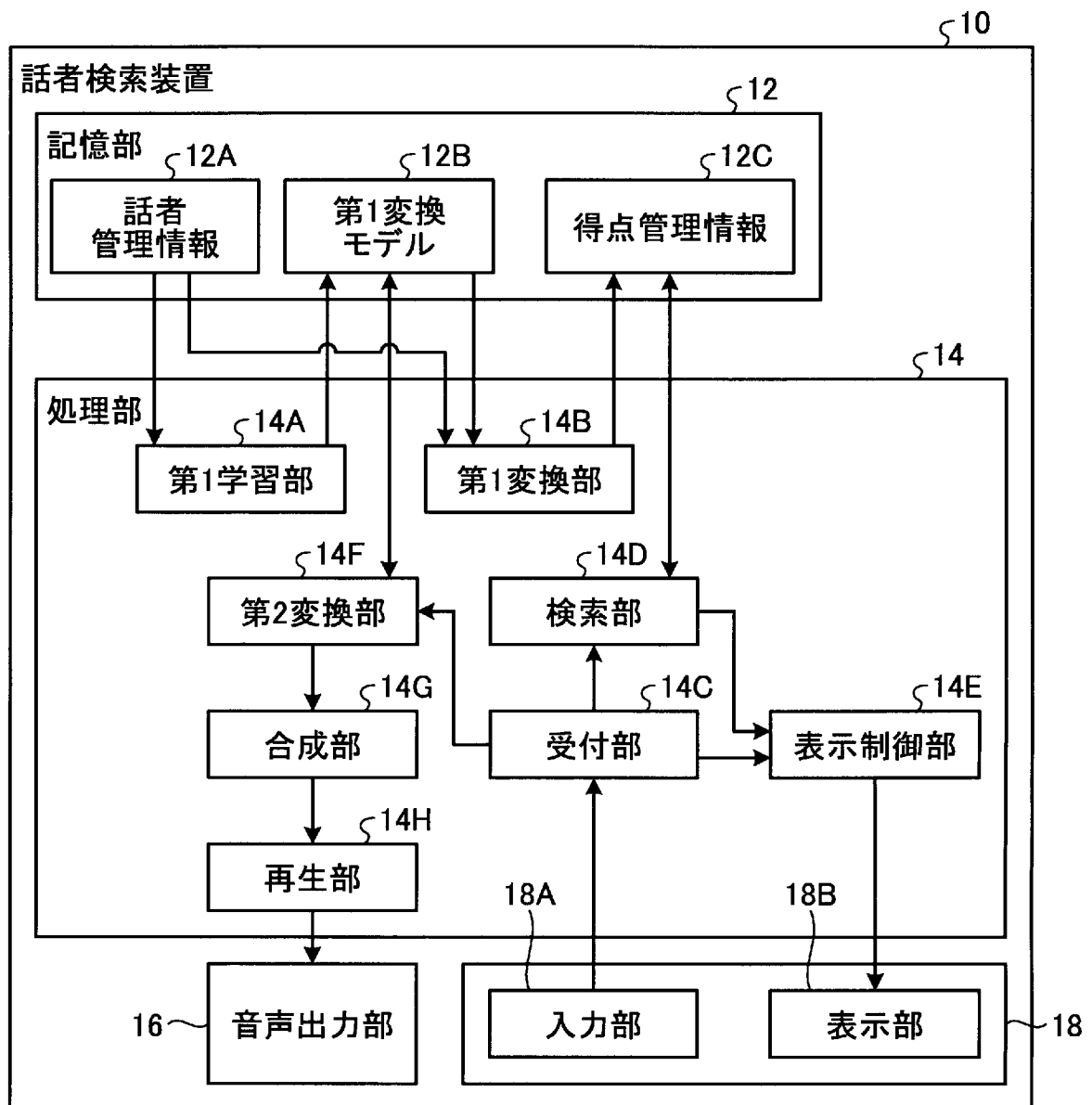
声質の特徴を表す得点ベクトルを音響モデルに変換するための第 1 変換モデルの逆変換モデルを用いて、予め登録された前記音響モデルを前記得点ベクトルに変換し、話者識別情報に対応づけて得点管理情報に登録するステップと、

前記得点ベクトルの入力を受付けるステップと、

前記得点管理情報から、受付けた前記得点ベクトルに類似する話者識別情報を検索するステップと、

をコンピュータに実行させるための話者検索プログラム。

[図1]



[図2]

12A

話者管理情報							得点ベクトル
話者ID	音声データ	音声のテキスト	音響特徴量	言語特徴量	音響モデル		
A							Sa
B							Sa
C							
D							
E							
F							Sa
G							Sa
H							Sa

[図3]

話者ID	声質の特徴						
	性別	年齢	明るさ	硬さ	明瞭さ	流暢さ	かすれ
A	0.5	0.1	1.0	0.2	0.7	0.1	0.2
B	0.1	0.2	0.8	0.4	0.9	0.8	0.1
C	0.9	0.5	0.4	0.5	0.9	0.8	0.3
D	1.0	0.9	0.2	0.6	0.4	0.7	0.9
...							

[図4]

12C

得点管理情報								得点ベクトル
話者ID	音声データ	音声のテキスト	音響特徴量	言語特徴量	音響モデル			
A								Sa
B								Sa
C								Sb
D								
E								Sb
F								Sa
G								Sa
H								Sa

[図5]

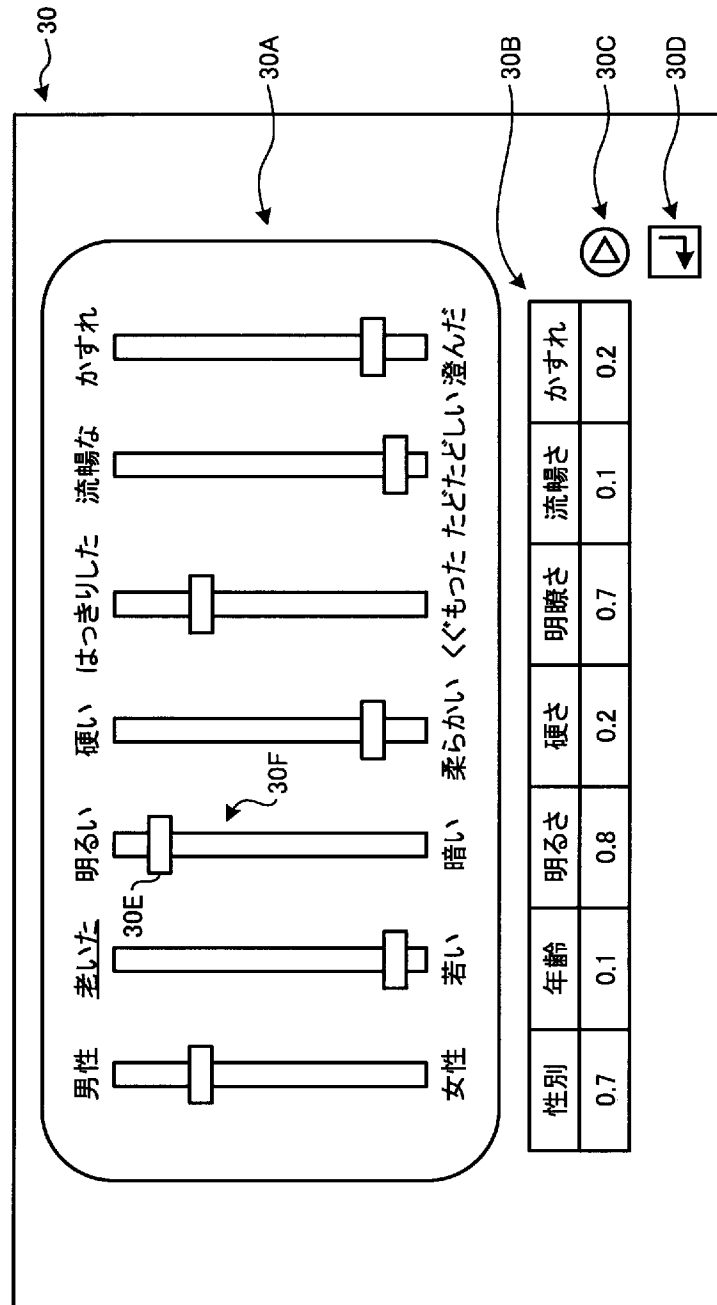


Figure 1 illustrates a system for generating a synthesized voice based on a target speaker's characteristics. The system includes a set of seven vertical sliders (30A-30G) and two tables (32B and 32C).

The sliders are labeled with gender (男性/Male, 女性/Female) and five attributes: 年齢 (Age), 明るさ (Brightness), 硬さ (Hardness), 明瞭さ (Clarity), and 流暢さ (Flow). Slider 30F is highlighted with a red arrow.

Below the sliders are two tables:

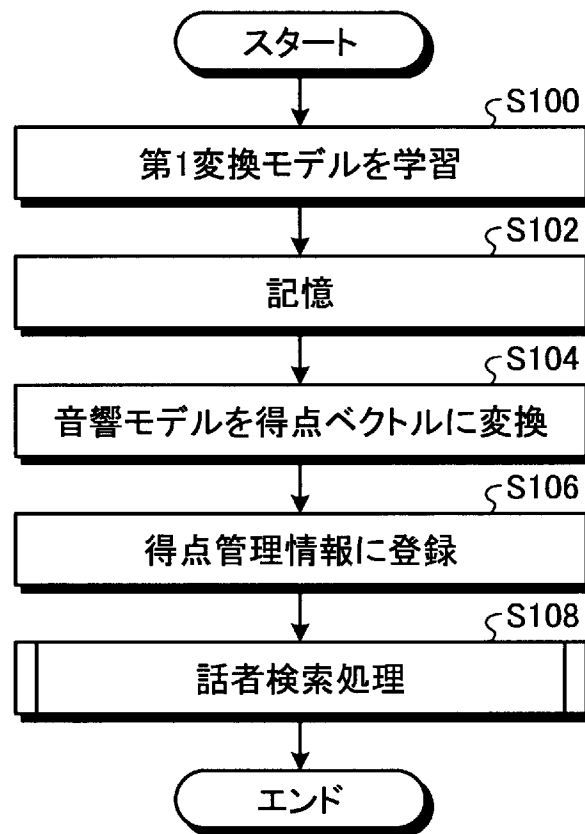
Table 1 (32B) shows numerical values for each attribute for a specific speaker (Speaker A):

性別	年齢	明るさ	硬さ	明瞭さ	流暢さ	かすれ
0.7	0.1	0.8	0.2	0.7	0.1	0.2

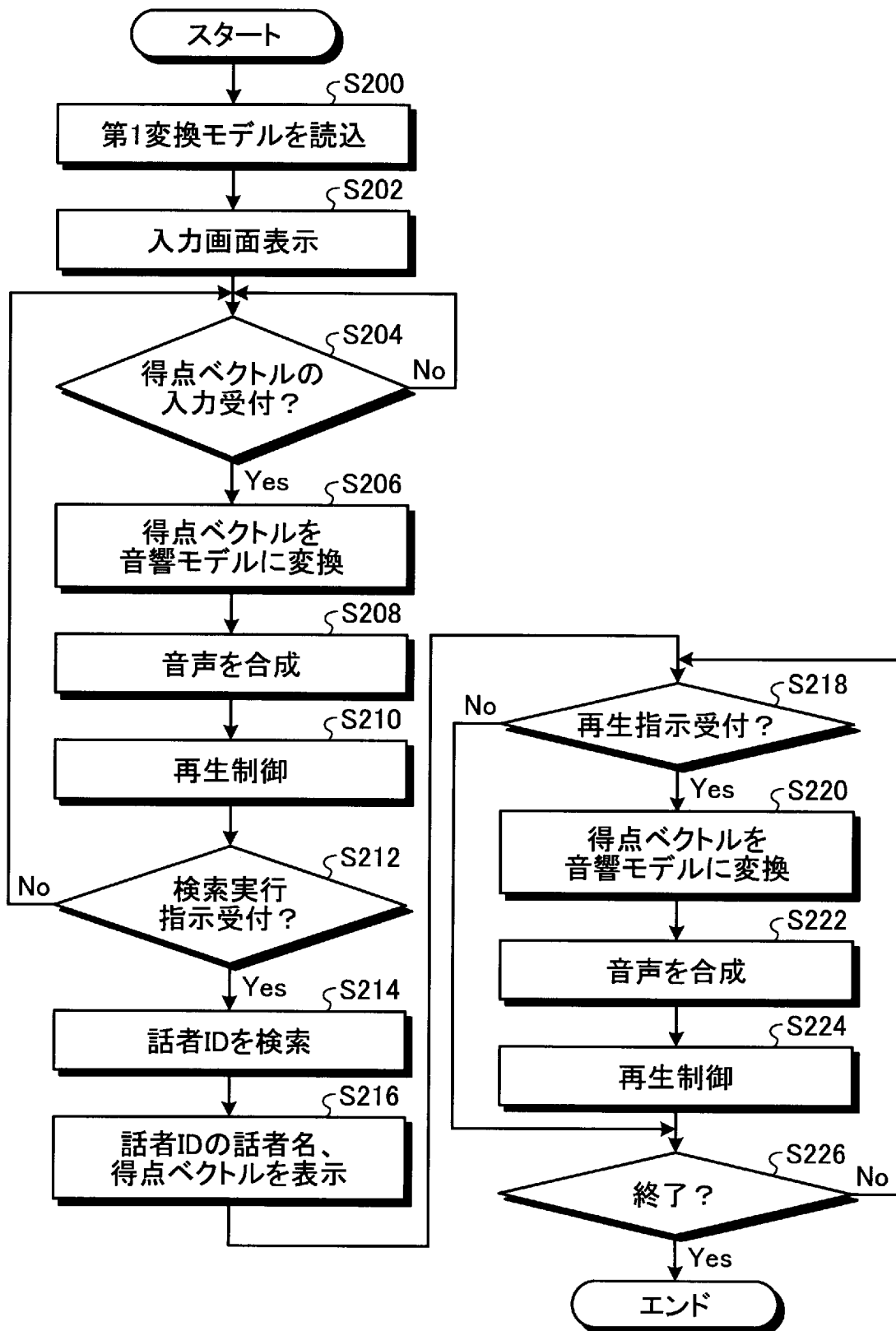
Table 2 (32C) shows numerical values for each attribute for three different speakers (Speakers A, F, and G):

性別	年齢	明るさ	硬さ	明瞭さ	流暢さ	かすれ
話者A	0.5	0.1	1.0	0.2	0.1	0.2
話者F	0.6	0.2	0.9	0.4	0.2	0.2
話者G	0.5	0.0	0.7	0.1	0.0	0.1

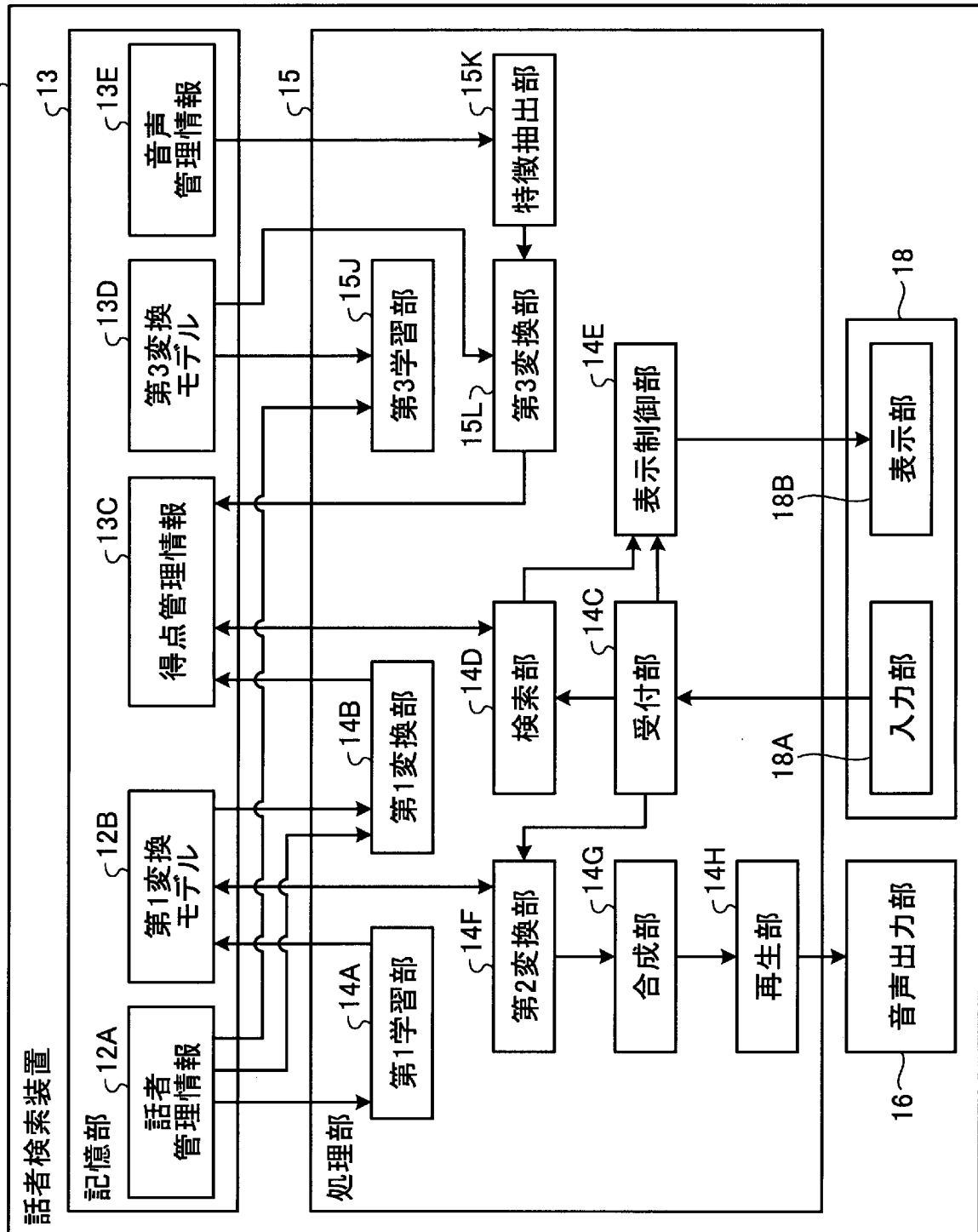
[図7]



[図8]



510A



[図10]

13E
↙

音声管理情報

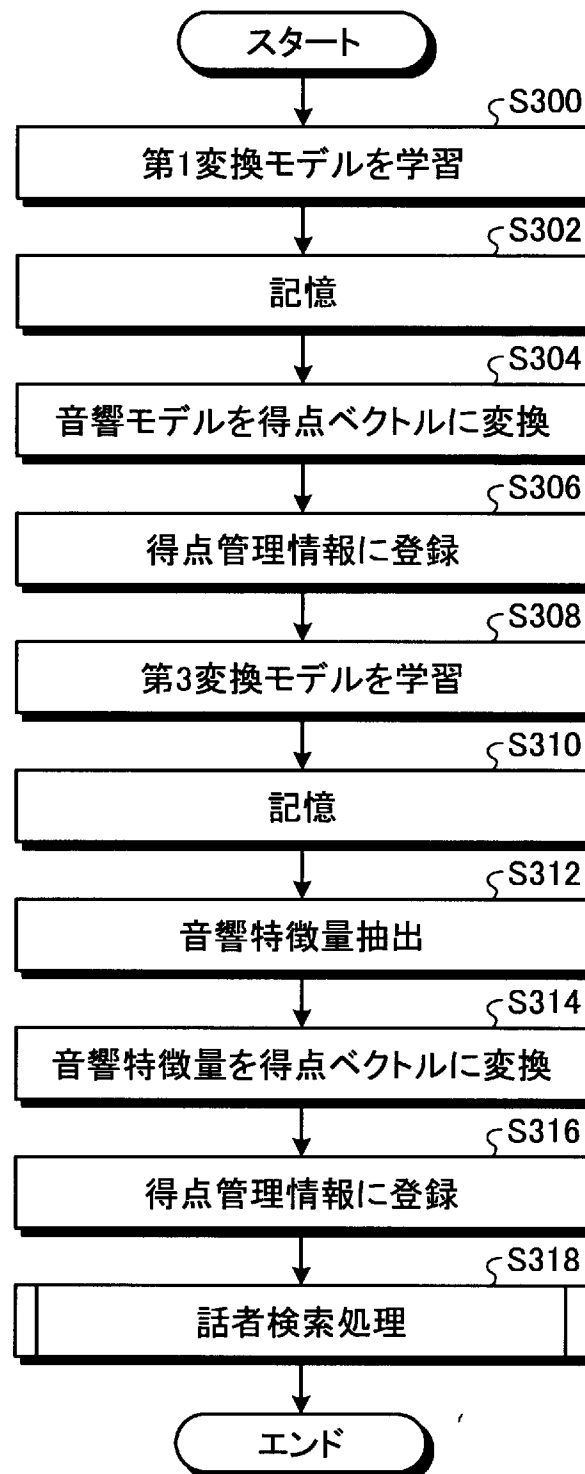
話者ID	音声データ
D	
X	
Y	
Z	

[図11]

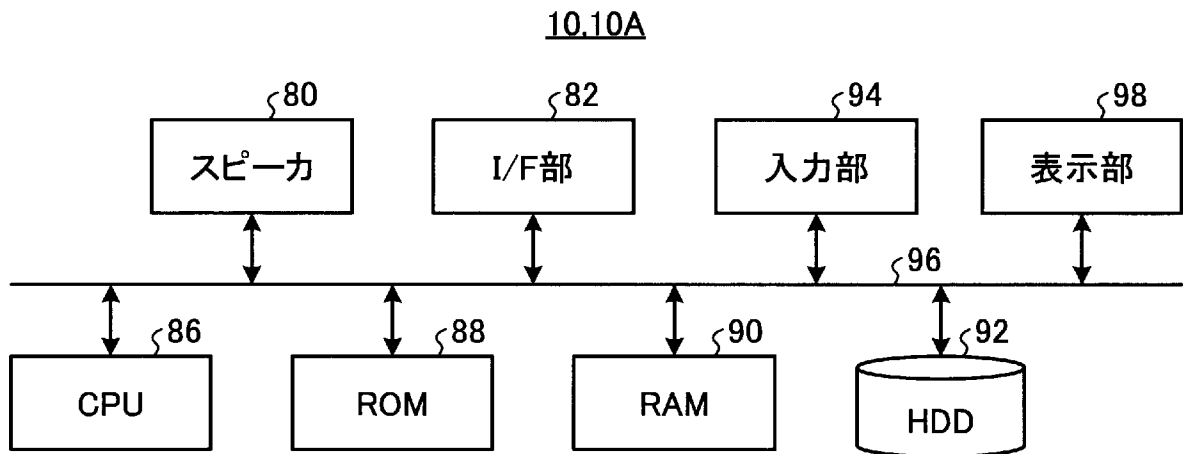
13C

得点管理情報							得点ベクトル
話者ID	音声データ	音声のテキスト	音響特徴量	言語特徴量	音響モデル		
A							Sa
B							Sa
C							Sb
D							Sc
E							Sb
F							Sa
G							Sa
H							Sa
X							Sc
Y							Sc
Z							Sc

[図12]



[図13]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2017/034573

A. CLASSIFICATION OF SUBJECT MATTER

Int. Cl. G06F17/30 (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)

Int. Cl. G06F17/30, G10L13/00, G10L15/00, G10L17/00

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 1971-2017

Registered utility model specifications of Japan 1996-2017

Published registered utility model applications of Japan 1994-2017

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 2011-70192 A (RICOH CO., LTD.) 07 April 2011, entire text & US 2011/0071833 A1, entire text & EP 2306345 A2 & CN 102023995 A	1-9
A	JP 2011-59703 A (SONY CORP.) 24 March 2011, entire text (Family: none)	1-9
A	JP 2015-172769 A (TOSHIBA CORP.) 01 October 2015, entire text & US 2013/0262119 A1, entire text & EP 2650874 A1 & CN 103366733 A	1-9
A	JP 2009-210790 A (NEC SOFTWARE KYUSHU, LTD.) 17 September 2009, entire text (Family: none)	1-9



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
16.11.2017

Date of mailing of the international search report
28.11.2017

Name and mailing address of the ISA/
Japan Patent Office
3-4-3, Kasumigaseki, Chiyoda-ku,
Tokyo 100-8915, Japan

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2017/034573

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 11-85181 A (NIPPON TELEGRAPH AND TELEPHONE CORP.) 30 March 1999, entire text (Family: none)	1-9

A. 発明の属する分野の分類 (国際特許分類 (IPC))

Int.Cl. G06F17/30(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. G06F17/30, G10L13/00, G10L15/00, G10L17/00

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	1971-2017年
日本国実用新案登録公報	1996-2017年
日本国登録実用新案公報	1994-2017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 2011-70192 A (株式会社リコー) 2011.04.07, 全文 & US 2011/0071833 A1, 全文 & EP 2306345 A2 & CN 102023995 A	1-9
A	JP 2011-59703 A (ソニー株式会社) 2011.03.24, 全文 (ファミリーなし)	1-9

☒ C欄の続きにも文献が列挙されている。

☐ パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー

「A」特に関連のある文献ではなく、一般的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

16.11.2017

国際調査報告の発送日

28.11.2017

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)

石田 信行

5M

9469

電話番号 03-3581-1101 内線 3599

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
A	JP 2015-172769 A (株式会社東芝) 2015.10.01, 全文 & US 2013/0262119 A1, 全文 & EP 2650874 A1 & CN 103366733 A	1-9
A	JP 2009-210790 A (九州日本電気ソフトウェア株式会社) 2009.09.17, 全文 (ファミリーなし)	1-9
A	JP 11-85181 A (日本電信電話株式会社) 1999.03.30, 全文 (ファミリーなし)	1-9