(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
2024年6月13日(13.06.2024)

WIPO | PCT

(10) 国際公開番号
WO 2024/122304 A1

(51) 国際特許分類:
G10L 17/00 (2013.01) *G10L 17/08* (2013.01)
G10L 15/04 (2013.01)

(21) 国際出願番号： PCT/JP2023/041303

(22) 国際出願日： 2023年11月16日(16.11.2023)

(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 2022-194509 2022年12月5日(05.12.2022) JP

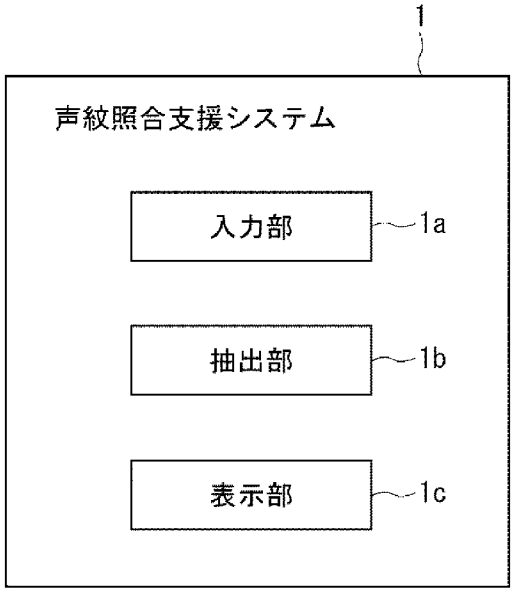
(71) 出願人: 日本電気株式会社 (**NEC CORPORATION**) [JP/JP]; 〒1088001 東京都港区芝五丁目７番１号 Tokyo (JP).

(72) 発明者: 後藤 晃 (**GOTOH Akira**); 〒1088001 東京都港区芝五丁目７番１号 日本電気株式会社内 Tokyo (JP). 古明地 秀治 (**KOMEIJI Shuji**); 〒1088001 東京都港区芝五丁目７番１号 日本電気株式会社内 Tokyo (JP). 中西 裕子 (**NAKANISHI Yuko**); 〒1088001 東京都港区芝五丁目７番１号 日本電気株式会社内 Tokyo (JP). 久我 優香 (**KUGA Yuka**); 〒1088001 東京都港区芝五丁目７番１号 日本電気株式会社内 Tokyo (JP).

(74) 代理人: 家入 健 (**IEIRI Takeshi**); 〒2210835 神奈川県横浜市神奈川区鶴屋町三丁目３３番８ アーバンセンター横浜ウエスト５階 響国際特許事務所 Kanagawa (JP).

(54) **Title:** VOICEPRINT COLLATION ASSISTANCE SYSTEM, VOICEPRINT COLLATION ASSISTANCE METHOD, AND PROGRAM

(54) 発明の名称： 声紋照合支援システム、声紋照合支援方法、及びプログラム



1 Voiceprint collation assistance system
1a Input unit
1b Extraction unit
1c Display unit

Fig. 1

(57) **Abstract:** Provided is a voiceprint collation assistance system that can suppress inclusion of a portion of voice data to be voiceprint-collated that is uttered by someone other than a subject of voiceprint collation when a portion uttered by the subject is presented to a user. The voiceprint collation assistance system pertaining to the present disclosure comprises an input unit, an extraction unit, and a display unit. The input unit inputs sample voice data of a subject for voiceprint collation. The extraction unit extracts a time segment in which the subject is speaking from voice data to be voiceprint-collated, on

WO 2024/122304 A1

［続葉有］

(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, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, 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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, 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, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, 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, ME, 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))

the basis of the sample voice data. The display unit presents the time segment extracted by the extraction unit to the user.

(57) 要約: 声紋照合対象音声データから声紋照合の対象者が発話している部分をユーザに提示するに際し、対象者以外が発話している部分が含まれることを抑制することが可能な声紋照合支援システムを提供する。本開示に係る声紋照合支援システムは、入力部、抽出部、及び表示部を備える。入力部は、声紋照合の対象者のサンプル音声データを入力する。抽出部は、サンプル音声データに基づき、声紋照合対象音声データから対象者が発話している時間区間を抽出する。表示部は、抽出部で抽出された時間区間をユーザに提示する。

明　細　書

発明の名称 ：

声紋照合支援システム、声紋照合支援方法、及びプログラム

技術分野

[0001]　本開示は、声紋照合支援システム、声紋照合支援方法、及びプログラムに関する。

背景技術

[0002]　声紋鑑定等のための声紋照合は、照合対象の話者が発声しているであろう区間を見つけるべく、作業者が音声全体又は音声を含む映像全体を検聴して手作業で抽出している。

[0003]　また、特許文献１には、話者を検索するシステムにおいて、検出された音声の話者が類似の場合、検出結果が本当に探したい人物が発したものであるかを判定することに困難を伴うといった課題を解決するための話者検索システムが開示されている。特許文献１に記載の話者検索システムは、音声データを蓄積する音声データベース、最適聴取区間検出部、話者検索部、及び検索結果提示部を備える。上記最適聴取区間検出部は、蓄積された音声データから話者の固有性が高い最適聴取区間を検出する。上記話者検索部は、ユーザが入力した音声または話者名に従い、当該音声または話者と同一話者が発話している音声データを蓄積された音声データから検索する。上記検索結果提示部は、上記話者検索部により得られた音声データに関する情報を、上記最適聴取区間検出部で検出された当該音声データの話者の固有性が高い最適聴取区間に関する情報とともに提示する。

先行技術文献

特許文献

[0004] 特許文献1：国際公開第２０１４／１５５６５２号

発明の概要

発明が解決しようとする課題

[0005] 上述のように、声紋鑑定等のための声紋照合は作業者による手作業が必要となるが、この作業は音声全体を繰り返し検聴する必要があるため、多大な時間を要し、疲労感により声紋照合の精度が落ちることや精度のばらつきも出てくることが容易に想像できる。

[0006] また、特許文献１には、同一話者の発話区間をクラスタリングにより検出する記載はあるものの、特許文献１に記載の技術では、発話区間内に特定話者以外の発話が重なることがあることを考慮したものではない。よって、特許文献１に記載の技術では、発話区間内に特定話者以外の発話が重なることを排除した状態で特定話者が発話している区間をユーザに提示することができず、提示する情報に改善の余地がある。

[0007] 本開示は、上記の課題を解決するためになされたもので、次のような声紋照合支援システム、声紋照合支援方法、及びプログラムを提供することをその目的とする。即ち、本開示は、声紋照合対象音声データから声紋照合の対象者が発話している部分をユーザに提示するに際し、対象者以外が発話している部分が含まれることを抑制することが可能な声紋照合支援システム等を提供することを、その目的とする。

課題を解決するための手段

[0008] 本開示に係る声紋照合支援システムは、声紋照合の対象者のサンプル音声データを入力する入力部と、前記サンプル音声データに基づき、声紋照合対象音声データから前記対象者が発話している時間区間を抽出する抽出部と、前記抽出部で抽出された前記時間区間をユーザに提示する表示部と、を備えたものである。

[0009] また、本開示に係る声紋照合支援方法は、声紋照合の対象者のサンプル音声データを入力する入力処理と、前記サンプル音声データに基づき、声紋照合対象音声データから前記対象者が発話している時間区間を抽出する抽出処理と、抽出した前記時間区間をユーザに提示するために表示装置に表示させる表示処理と、を備えたものである。

[0010] また、本開示に係るプログラムは、コンピュータに、声紋照合の対象者の

サンプル音声データを入力する入力処理と、前記サンプル音声データに基づき、声紋照合対象音声データから前記対象者が発話している時間区間を抽出する抽出処理と、抽出した前記時間区間をユーザに提示するために表示装置に表示させる表示処理と、を含む声紋照合支援処理を実行させるプログラムである。

発明の効果

[0011] 本開示によれば、声紋照合対象音声データから声紋照合の対象者が発話している部分をユーザに提示するに際し、対象者以外が発話している部分が含まれることを抑制することが可能な声紋照合支援システム等を提供することができる。

図面の簡単な説明

[0012] [図1]本開示に係る声紋照合支援システムの一構成例を示すブロック図である。

[図2]図１の声紋照合支援システムにおける声紋照合支援方法の一例を説明するためのフロー図である。

[図3]本開示に係る声紋照合支援システムの他の構成例を示すブロック図である。

[図4]図３の声紋照合支援システムで処理対象となる声紋照合対象音声データの波形の一例を示す図である。

[図5]図３の声紋照合支援システムで処理対象となる声紋照合対象音声データの一例を概念的に示す概念図で、図４の声紋照合対象音声データの波形を、話者毎に分けて模式的に示す図である。

[図6]図３の声紋照合支援システムで使用されるサンプル音声データの波形の一例を示す図である。

[図7]図３の声紋照合支援システムで使用されるサンプル音声データの一例を概念的に示す概念図で、図６のサンプル音声データの波形を模式的に示す図である。

[図8]図４の声紋照合対象音声データの波形に対し、抽出された時間区間を表

示する表示例を示す図である。

[図9]図４の声紋照合対象音声データの波形に対し、抽出された時間区間を表示する他の表示例を示す図である。

[図10]図４の声紋照合対象音声データの波形に対し、抽出された時間区間を表示する他の表示例を示す図である。

[図11]図４の声紋照合対象音声データの波形に対し、抽出された時間区間を表示する他の表示例を示す図である。

[図12]比較例１に係る声紋照合対象ファイルの作成作業を説明するためのフロー図である。

[図13]比較例２に係る声紋照合対象ファイルの作成作業を説明するためのフロー図である。

[図14]装置に含まれるハードウェア構成例を示す図である。

発明を実施するための形態

[0013] 以下、実施形態について、図面を参照しながら説明する。説明の明確化のため、以下の記載及び図面は、適宜、省略、及び簡略化がなされている。また、各図面において、同一の要素には同一の符号が付されており、必要に応じて重複説明は省略されている。

[0014] ＜実施形態１＞

図１は、本開示に係る声紋照合支援システムの一構成例を示すブロック図である。図１に示す本開示に係る声紋照合支援システム１は、声紋照合を行う対象の候補となる音声データをユーザに提示することで、ユーザが声紋照合を行う作業の支援を行うシステムとすることができる。声紋照合支援システム１は、例えば警察の鑑識などでの声紋鑑定作業や、その他の声紋照合が必要な様々な作業において利用することができる。

[0015] このような支援のために、図１に示す声紋照合支援システム１は、入力部１ａ、抽出部１ｂ、及び表示部１ｃを備えることができる。

[0016] 入力部１ａは、声紋照合の対象者のサンプル音声データを入力する。ここで、声紋照合の対象者は１名とする。サンプル音声データは、声紋対象の対

象者のみの発話を含むように録音された１チャンネル又は複数チャンネルの音声データとすることができる。サンプル音声データの入力は、抽出部１ｂでの抽出の実行時に参照されるため、抽出の実行前に参照可能に声紋照合支援システム１の内部に記憶しておけばよい。入力元は、例えば音声データを取得するマイクロフォン等の音声取得装置、あるいは、予め取得されたサンプル音声データを記憶したサーバコンピュータ等のコンピュータなどが挙げられる。これらの音声取得装置あるいはコンピュータは声紋照合支援システム１に備えることもできる。入力部１ａは、サンプル音声データを入力し、抽出部１ｂに渡す部位であり、そのための通信インタフェース等の入力インタフェースを備えることができる。

[0017] 抽出部１ｂは、サンプル音声データに基づき、声紋照合対象音声データから対象者が発話している時間区間を抽出する。この抽出は、サンプル音声データに基づくものであるため、抽出部１ｂでは、両データが分析され、サンプル音声データに関連性が高い時間区間が抽出されることになる。抽出部１ｂは、例えば、サンプル音声データを所定の時間間隔で区切り、声紋照合対象音声データも上記所定の時間間隔で区切り、上記所定の時間間隔毎に、対象者が発話している時間区間を抽出することができる。なお、ここでは、声紋照合対象音声データの音声内容を認識する処理は不要である。

[0018] また、抽出部１ｂは、例えば、サンプル音声データと声紋照合対象音声データとを入力し、対象者が発話している時間区間を出力するように学習された学習モデルを用いて、時間区間の抽出を行うこともできる。この学習モデルのアルゴリズム等は問わず、また入力対象のデータは例えば上記所定の時間間隔で区切ったものとすることができる。

[0019] 声紋照合対象音声データは、録音された１チャンネル又は複数チャンネルの音声データとすることができる。特に、本実施形態では、対象者のサンプル音声データに基づき声紋照合対象音声データを抽出するため、声紋照合対象音声データが喩えモノラルの音声データであっても、対象者が発話している時間区間を抽出することは可能である。また、声紋照合対象音声データは

、映像データに関連付けられたデータとすること、換言すれば音声付き動画データとすることもできる。

[0020] また、声紋照合対象音声データは、予めサーバコンピュータ等のコンピュータに記憶しておき、そのコンピュータから取得されること、あるいは声紋照合支援システム１に具備された記憶装置に記憶しておき、その記憶装置から読み出すことができる。このコンピュータは声紋照合支援システム１に備えることもできる。

[0021] あるいは、声紋照合対象音声データは、リアルタイムでなされている会話をマイクロフォン等の音声取得装置で取得したデータとすること、リアルタイムでなされている電話での通話内容を取得したデータとすることもできる。これらの例のように声紋照合対象音声データがリアルタイムで取得されるデータである場合、抽出や表示部１ｃによる提示の処理はリアルタイムに行うこともできる。

[0022] 表示部１ｃは、抽出部１ｂで抽出された時間区間をユーザに提示する。表示部１ｃは、表示装置と、表示装置での表示の制御を行う表示制御部と、を備えることができる。この表示装置は、画像を表示できればよく、ＬＣＤ（Liquid Crystal Display）、有機ＥＬ（Electro-Luminescence）ディスプレイ等のディスプレイとすることや、プロジェクタとすることができる。表示装置は、例えば、スマートフォン又はタブレット端末等に含まれるディスプレイであってもよい。

[0023] また、声紋照合支援システム１は、例えば、サーバコンピュータ又はパーソナルコンピュータ又はスマートフォン等のコンピュータとすることができる。具体的には、声紋照合支援システム１は、例えば１以上のプロセッサと１以上のメモリとを含むハードウェアを含むコンピュータ装置を含んで構成され得る。声紋照合支援システム１内の各部の機能の少なくとも一部は、１以上のプロセッサが、１以上のメモリから読み出したプログラムに従って動作することで実現され得る。

[0024] 換言すれば、声紋照合支援システム１は、その全体を制御する制御部（図

示せず）を備えることができる。この制御部は、例えば、プロセッサ、作業用メモリ、及びプログラムを記憶した不揮発性の記憶装置などによって実現することができる。このプロセッサは、例えばＣＰＵ（Central Processing Unit）又はＧＰＵ（Graphics Processing Unit）等とすることができる。このプログラムは、入力部１ａの入力処理、抽出部１ｂの抽出処理、及び表示部１ｃの表示制御の処理をプロセッサに実行させるためのプログラムとすることができる。また、声紋照合支援システム１は、入力されたサンプル音声データ、声紋照合対象音声データ、抽出された時間区間を示す情報などを記憶する記憶装置を備えることができ、この記憶装置は例えば上記の制御部に備えられる記憶装置を利用することもできる。なお、声紋照合支援システム１は、専用のハードウェアを備えた装置として構成することもできる。

[0025] また、声紋照合支援システム１は、単体の装置として構成される例に限らず、機能を分散させた複数の装置として構築することもでき、その分散の方法は問わない。分散させる装置の一つには無論、表示装置を含むことができる。複数の装置に機能を分散した声紋照合支援システムを構築する場合、各装置に制御部、通信部、及び必要に応じて記憶部等を備える。さらに、この場合、無線又は有線の通信により上記複数の装置を必要に応じて接続して協働して声紋照合支援システム１で説明した機能を実現させればよい。

[0026] 次に、図２を参照しながら、声紋照合支援システム１の処理例について説明する。図２は、図１の声紋照合支援システム１における声紋照合支援方法の一例を説明するためのフロー図である。

[0027] この声紋照合支援方法は、次のような声紋照合支援処理を実行する。声紋照合支援処理では、声紋照合の対象者のサンプル音声データを入力する入力処理を実行する（ステップＳ１）。次いで、声紋照合支援処理では、そのサンプル音声データに基づき、声紋照合対象音声データから対象者が発話している時間区間を抽出する抽出処理を実行する（ステップＳ２）。そして、声紋照合支援処理では、抽出した時間区間をユーザに提示するために表示装置に表示させる表示処理を実行し（ステップＳ３）、処理を終了する。なお、

上記のプログラムは、コンピュータに、このような入力処理、抽出処理、及び表示処理を含む声紋照合支援処理を実行させるプログラムとすることができる。

[0028] このように、声紋照合支援システム１は、声紋照合の対象者（話者）のサンプル音声データを入力し、声紋照合対象音声データ中の、サンプル音声データと関連性の高い時間区間を分析して抽出し、ユーザにこの時間区間を示す時間情報を戻す。

[0029] このような声紋照合支援処理の後、提示を受けたユーザは、提示内容を考慮して、声紋照合対象音声データから声紋照合を行う音声データを選択し、検聴作業や声紋照合装置による声紋照合処理を行うことができる。本実施形態では、ユーザはこのような声紋照合の支援を受けて、よりユーザの検聴作業の負荷や声紋照合装置の処理負荷がかかるような声紋照合処理を、選択した音声データに対して実行することができる。なお、この音声データが実際に声紋照合を行う対象のデータとなり、一方で、抽出部１ｂで処理対象となる上記声紋照合対象音声データは声紋照合の対象を抽出するための音声データである。よって、上記声紋照合対象音声データは、抽出対象データあるいは選択対象データなどと称することもできる。

[0030] 以上に説明したように、本実施形態によれば、対象者以外が発話している部分が含まれることを抑制して、声紋照合対象音声データにおける声紋照合の対象者が発話している部分を、ユーザに提示することが可能になる。

[0031] この効果について補足する。声紋照合、つまり音声に基づく話者照合を高精度に行うためには、特定話者だけが発話している区間を抽出する必要がある。現状、声紋鑑定等の声紋照合作業においては、長時間の音声データを検聴して声紋照合に適した時間区間を抽出しているため、作業者の長時間作業による負担が大きく且つ作業者の主観が入り精度が安定しないことが懸念される。これに対し、本実施形態では、特定話者として声紋照合の対象者だけが発話している時間区間、特に他者の発話との重なりがない時間区間を抽出することを目的として、その対象者のサンプル音声データに基づき抽出を実

行し、ユーザにその時間区間を提示している。その対象者のサンプル音声データに基づく抽出となるため、他者の発話との重なりが少ないような抽出結果が得られる。そして、ユーザは、声紋照合支援システム１において対象者が発話している時間区間を自動的に抽出した結果を確認することで、検聴等により声紋照合を行う時間区間の絞り込みを行うことができる。

[0032] このように、本実施形態では、発話区間内に特定話者以外の発話が重なることを抑制した状態で特定話者が発話している区間をユーザに提示することができる。よって、本実施形態によれば、声紋鑑定等の声紋照合が必要な作業において、作業者となるユーザは提案された区間を中心に検聴することで、一部始終検聴（視聴）していた作業を効率化することができる。そのため、本実施形態によれば、ユーザの抽出にかかる手作業についての手間の低減、精度低下の防止、精度ばらつきの防止を行うことができる。

[0033] また、本実施形態では、声紋照合の対象者が１名であることを前提として説明したが、複数名であってもよい。その場合、その複数名のそれぞれのサンプル音声データに基づき、声紋照合対象音声データの中からその複数名が同時に会話している時間区間を抽出し、ユーザに提示することができる。

[0034] ＜実施形態２＞

実施形態２について、図３～図１３を参照しながら実施形態１との相違点を中心に説明するが、実施形態１で説明した様々な例が適用できる。まず、図３を参照しながら本開示に係る声紋照合支援システムの他の構成例について説明する。図３は、本開示に係る声紋照合支援システムの他の構成例を示すブロック図である。

[0035] 図３に示すように、本開示に係る声紋照合支援システム１００は、サーバコンピュータ（以下、サーバ）１０、ユーザ端末２０、処理対象音声取得装置３０、及びサンプル音声取得装置４０を備えることができる。

[0036] サーバ１０は、制御部１１、記憶部１２、及び通信部１３を備えることができる。制御部１１は、サーバ１０の全体を制御する部位で、例えばＣＰＵ又はＧＰＵ等のプロセッサ、作業用メモリ、及びプログラムを記憶した不揮

発性の記憶装置などによって実現することができる。このプログラムは、入力部１ａの入力処理、抽出部１ｂの抽出処理、及び表示部１ｃの表示制御の処理の一部を含む処理をプロセッサに実行させるためのプログラムとすることができる。ここで、表示制御処理の一部とは、ユーザ端末２０において、抽出された時間区間を表示させるように、通信部１３を介してユーザ端末２０へ指示する処理を指すことができる。制御部１１により、上記のプログラムが読み出されて実行されることで、サーバ１０は、入力部１ａの入力機能、抽出部１ｂの抽出機能、及び表示部１ｃにおける表示制御機能の一部を含む機能を実現することができる。

[0037] 記憶部１２は、サンプル音声取得装置４０から入力されたサンプル音声データ、処理対象音声取得装置３０から入力された声紋照合対象音声データ（以下、処理対象音声データ）、及び、抽出された時間区間を示す情報などを記憶する記憶装置とすることができる。記憶部１２は、抽出処理に使用する閾値や時間間隔も記憶しておくことができる。

[0038] 通信部１３は、ユーザ端末２０、処理対象音声取得装置３０、及びサンプル音声取得装置４０と、有線又は無線のネットワークを介して、通信する通信インタフェースとすることができる。

[0039] ユーザ端末２０は、パーソナルコンピュータ、スマートフォンなどの情報処理装置とすることができ、制御部２１、記憶部２２、通信部２３、及び表示部２４を備えることができる。

[0040] 制御部２１は、ユーザ端末２０の全体を制御する部位で、例えばＣＰＵ又はＧＰＵ等のプロセッサ、作業用メモリ、及びプログラムを記憶した不揮発性の記憶装置などによって実現することができる。このプログラムは、通信部２３を介してサーバ１０にアクセスする処理、及び表示部１ｃの表示制御の処理の一部を含む処理をプロセッサに実行させるためのプログラムとすることができる。制御部２１により、上記のプログラムが読み出されて実行されることで、ユーザ端末２０は、これらの処理を行う機能を実現することができる。ここで、表示制御機能の一部とは、抽出された時間区間を表示させ

る指示をサーバ１０から通信部２３を介して受信した場合に、その指示に従い表示部２４に表示させる処理を指すことができる。

[0041] 記憶部２２は、時間区間の表示に関する様々な設定等を記憶する記憶装置とすることができる。この設定には、例えば、時間区間を表示するためのユーザインタフェースでの表示設定を含むことができる。通信部２３は、サーバ１０と、有線又は無線のネットワークを介して通信する通信インタフェースとすることができる。

[0042] 処理対象音声取得装置３０は、サーバ１０で処理対象となる処理対象音声データを取得し、サーバ１０へ送信する装置である。処理対象音声取得装置３０は、電話システム、音声通話が可能なネットワークシステム、マイクロフォン等を備えた録音装置などとすることができる。

[0043] サンプル音声取得装置４０は、サーバ１０で抽出処理に使用するサンプル音声データを取得し、サーバ１０へ送信する装置であり、マイクロフォン等を備えた録音装置などとすることができる。

[0044] 上述のような構成により、声紋照合支援システム１００は、処理対象音声取得装置３０で取得された処理対象音声データをサーバ１０に記憶すること、並びに、サンプル音声取得装置４０で取得されたサンプル音声データをサーバ１０に記憶することができる。この記憶の順序は問わない。

[0045] そして、サーバ１０の制御部１１は、抽出部１ｂの機能として説明したように、記憶されたサンプル音声データに基づき、記憶された処理対象音声データから対象者が発話している時間区間を抽出する抽出処理を実行する。

[0046] ここで、実施形態１での抽出処理は、対象者のサンプル音声データに基づき実行されるため、処理対象音声データから抽出される抽出後音声データには、対象者が発話している区間の音声データを含み、その区間がユーザに提示される。しかしながら、実施形態１での抽出処理では、対象者が他者と会話している区間の音声データも抽出されることがある。

[0047] そのため、本実施形態での抽出処理は、時間区間として、対象者が単独で発話している時間区間を抽出すること、つまり他者が発話している区間を排

除して時間区間を抽出する。

[0048] また、制御部１１は、抽出した時間区間をユーザに提示するためにユーザ端末２０で表示させる指示を、通信部１３を介してユーザ端末２０に送信する。ユーザ端末２０では、この指示を通信部２３でサーバ１０から受信し、制御部２１がその時間区間を表示部２４に表示させる制御を行い、表示部２４での時間区間の表示がなされる。

[0049] これにより、ユーザは表示として提案された時間区間を中心に検聴することで、一部始終検聴（視聴）していた作業を効率化できるようになる。なお、表示させる時間区間を示す情報はあくまで提案であり、この時間区間についての声紋照合を自動で行うものではない。無論、例えば、この時間区間についての声紋照合のうち、ユーザ端末２０において、声紋照合が必要な時間区間をユーザが選択し、声紋照合を図示しない声紋照合装置で実行させることもできる。この声紋照合装置はサーバ１０に搭載することもできる。

[0050] 本実施形態では、上述のような抽出処理を採用することにより、対象者が他者と会話しており且つ他者の発話が混じる区間の音声データを抽出対象外とすることができ、実施形態１と比較して、他者の声の重なりが無くなる。つまり、本実施形態によれば、声紋照合時に他者の音声が混じることを防ぐことができる。

[0051] このように、本実施形態によれば、処理対象音声データから声紋照合の対象者が発話している部分をユーザに提示するに際し、対象者以外が発話している部分が含まれることを排除することができる。換言すれば、本実施形態では、発話区間内に特定話者以外の発話が重なることを排除した状態で特定話者が発話している区間をユーザに提示することができる。

[0052] よって、本実施形態によれば、作業者となるユーザは提案された区間を中心に検聴することで、一部始終検聴（視聴）していた作業を、実施形態１に比してさらに効率化することができる。

[0053] また、制御部１１は、抽出処理として、サンプル音声データの特徴量と処理対象音声データの特徴量との類似度スコアに基づき、時間区間を抽出する

ようにすることができる。特徴量の種類、特徴量の算出、類似度スコアの算出方法は既存の手法を利用すればよい。特徴量は、例えば所定の時間間隔毎に算出され、類似度スコアもその所定の時間間隔毎に算出され、時間区間もその所定の時間間隔を単位として抽出されることができる。この手法では、処理対象音声データにおける所定の時間間隔の区間のうち、サンプル音声データの特徴量が示す対象者とは異なる他者の音声が混じっている区間では、類似度スコアが下がることになる。

[0054] 特に、制御部１１は、抽出処理において、時間区間として、処理対象音声データにおける類似度スコアが所定閾値以上に該当する区間を抽出するようにしてもよい。このような抽出処理を採用することで、他者の音声が混じる区間を排除した状態で、ユーザに時間区間を提示することができる。これにより、ユーザに対象者と関連性が高い（類似度スコアが高い）情報のみを、つまり検聴に必要な区間を示す情報のみを見易く提示することができる。

[0055] このような処理の具体例について、図４～図１１を参照しながら説明する。図４は、声紋照合支援システム１００で処理対象となる処理対象音声データの波形の一例を示す図である。また、図５は、声紋照合支援システム１００で処理対象となる処理対象音声データの一例を概念的に示す概念図で、図４の処理対象音声データの波形を、話者毎に分けて模式的に示す図である。図６は、声紋照合支援システム１００で使用されるサンプル音声データの波形の一例を示す図である。また、図７は、声紋照合支援システム１００で使用されるサンプル音声データの一例を概念的に示す概念図で、図６のサンプル音声データの波形を模式的に示す図である。

[0056] 処理対象音声取得装置３０で取得されサーバ１０で処理対象となる処理対象音声データは、例えば図４に示すような波形となる。図４で例示する波形の処理対象音声データは、例えば周波数分析やクラスタリングなどにより、特徴量に応じて１又は複数の話者や１又は複数の騒音に分けることができる。図５にはその一例として、話者Ａ、話者Ｂ、話者Ｃ、及び騒音についての特徴量を模式的に示した例を挙げている。なお、特徴量としては、音量、音

圧をはじめとする様々な種類の特徴量を示す値である。

[0057] また、サンプル音声取得装置４０で取得されサーバ１０での抽出処理に使用するサンプル音声データは、例えば図６に示すような波形となる。図６では、話者Ｃのサンプル音声データの例を挙げている。なお、図５との対比のため、図６で例示する波形のサンプル音声データの特徴量の例を模式的に示すと、図７のようなる。また、図５は複数の話者及び騒音（雑音）の重なり具合の様子の一例を説明するために図示したものに過ぎず、図７で示す話者Ｃの特徴量が抽出可能であることを説明するために図示したに過ぎない。

[0058] 時間区間の表示例について、図８～図１１を参照しながら説明する。図８は、図４の処理対象音声データの波形に対し、抽出された時間区間を表示する表示例を示す図である。また、図９～図１１は、図４の処理対象音声データの波形に対し、抽出された時間区間を表示する他の表示例を示す図である。

[0059] サーバ１０は、表示部２４に対し、図８で例示するように、処理対象音声データに対応する時間軸を含む処理対象音声データの波形を表示させた状態で、抽出された時間区間を、他の区間とは異なる表示形態で表示させるように指示することができる。そして、表示部２４はこのような表示を実行することができる。図８の例では、波形上に類似度スコアが所定閾値以上の算出対象時間を示す矩形領域ｃｓ１、ｃｓ２を示している。ここでは類似度スコアが０～１００の範囲の値として所定閾値を７０とした例を挙げている。図８において、スコア値の算出対象時間を示す矩形領域ｃｓ１、ｃｓ２はそれぞれ、図５における話者Ｃの特徴量のうちの特徴量ｃ１、ｃ２に対応する領域となる。

[0060] 表示形態はこれに限らず、抽出された時間区間をハイライトで表示させることもできるし、類似度スコアが高いほど濃い階調又は色の違いで表示させることもができる。後者の例は、類似度スコアを複数の閾値の範囲で分け、その範囲毎に異なる表示形態で表示させる例である。これにより、類似度スコアに基づき、関連性が高い時間区間をランキング形式で表示させることが

でき、これによりユーザが関連性の高さを視認し易くなる。また、ランキングした番号も併せて表示させることもできる。また、表示形態の他の例としては、図８の例において、図４の波形の代わりに図５で例示した特徴量を表示させることもできる。

[0061] また、時間軸を含む波形を表示した例を挙げたが、波形は表示させずに単に時間軸だけ表示させることもできる。例えば、図９で例示するように、時間軸上に類似度スコアが所定閾値以上の算出対象時間を示す矩形領域ｃｓ１、ｃｓ２を表示させることができる。図９において、類似度スコアの大きさ（値を示す高さ）及び類似度スコアの算出対象時間を示す矩形領域ｃｓ１、ｃｓ２はそれぞれ、図５における話者Ｃの特徴量のうちの特徴量ｃ１、ｃ２に対応する領域となる。

[0062] 図８の表示例は、類似度スコアの値も示しており、次のような処理にも該当する。即ち、サーバ１０は、表示部２４に対し、処理対象音声データに対応する時間軸を含む処理対象音声データの波形を表示させた状態で、少なくとも抽出された時間区間について、類似度スコアを可視化して表示させるように指示することもできる。そして、表示部２４はこのような表示を実行することができる。ここで類似度スコアの可視化とは、図８の例のように類似度スコアの値そのものを可視化してもよいし、図９の例のように類似度スコアの値が示す大きさが分かるように可視化してもよい。

[0063] サーバ１０は、表示部２４に対し、図１０で例示するように、処理対象音声データに対応する時間軸を含む処理対象音声データの波形又は時間軸のみを表示させた状態で、少なくとも抽出された時間区間について類似度スコアの疎密状態を表示させることもできる。つまり、サーバ１０は、表示部２４に対し、類似度スコアが高い区間の疎密状態を可視化して表示させるように指示することができる。類似度スコアが高い区間とは、例えば他の所定閾値以上に高い区間を指すことができる。上記他の所定閾値は上記所定閾値より低く設定しておくこともできる。そして、表示部２４はこのような表示を実行することができる。

[0064] 図１０の例では、疎密状態を、疎の場合にハッチング無し、密の場合にハッチング有りとし、密になるほどハッチングの濃さを上昇させた状態で表示させている。但し、図１０の例では、図８の例に比べて抽出単位となる上記所定の時間間隔が短い例を挙げている。疎密状態の表示はこの例に限らず、類似度スコアが高い部分が密集しているところが視認できるようなものであればよく、所謂一次元ヒートマップ又はそれに類似した表示形態で疎密状態を表現すればよい。

[0065] また、図８と図１０の違いからも分かるように、サーバ１０は、抽出処理での抽出単位となる時間間隔（上記所定の時間間隔）を設定する間隔設定部を備えることもできる。この間隔設定部は、例えば、かかる設定を実行する間隔設定プログラムを、制御部１１が実行可能な状態で記憶部１２に記憶させておくことで搭載することができる。これにより、ユーザは、時間間隔を変更して提案内容を変更しながら検聴する区間を選択することができる。

[0066] このような時間間隔の設定はユーザインタフェース（ＵＩ）で受け付けることができる。例えば、サーバ１０は、自発的に又はユーザ端末２０からの要求により、ユーザ端末２０に図１１で例示するようなＵＩ５０を提示し、表示部２４に表示させることができる。ユーザは、表示部２４に表示されたＵＩ５０において、抽出時間間隔の入力欄５５に間隔を入力することで、その下の領域に表示されている時間区間を示す情報を含むグラフ５８が、入力された時間間隔に基づき変更され、表示されることになる。

[0067] なお、この例ではグラフ５８において、図８の例が表示される例を挙げているが、図９又は図１０で例示したグラフを表示させるなど、これ以外のグラフを表示させることもでき、また、グラフ種の選択も可能としておくこともできる。

[0068] ＵＩ５０は、グラフ５８の他に入力領域５１を含み、この入力領域５１に入力欄５５を含めておくことができる。入力領域５１には、その他、例えば、低スコア領域削除ボタン５２、選択領域を残し削除ボタン５３、選択領域削除ボタン５４、抽出閾値の入力欄５６声紋照合対象ファイル作成ボタン５

７などを設けておくことができる。

[0069] また、サーバ１０は、上記の所定閾値を設定する閾値設定部を備えることもできる。この閾値設定部は、例えば、かかる設定を実行する閾値設定プログラムを、制御部１１が実行可能な状態で記憶部１２に記憶させておくことで搭載することができる。このような所定閾値の設定もＵＩで受け付けることができる。例えば、サーバ１０は、自発的に又はユーザ端末２０からの要求により、ユーザ端末２０に図１１で例示するような抽出閾値の入力欄５６を含むＵＩ５０を提示し、表示部２４に表示させることができる。ユーザは、ＵＩ５０において、抽出閾値の入力欄５６に間隔を入力することで、その下の領域に表示されているグラフ５８が、入力された抽出閾値に基づき変更され、表示されることになる。

[0070] また、サーバ１０は、処理対象音声データから、抽出された時間区間以外の区間のデータを削除する削除部を備えることができる。この削除部は、制御部１１の一機能として搭載することができ、例えば、かかる削除を実行する削除プログラムを、制御部１１が実行可能な状態で記憶させておくことで搭載することができる。このような削除もＵＩで受け付けることができる。

[0071] 例えば、サーバ１０は、自発的に又はユーザ端末２０からの要求により、ユーザ端末２０に図１１で例示するようなボタン５２～５４の少なくとも１つを含むＵＩ５０を提示し、表示部２４に表示させることができる。ユーザは、表示部２４に表示されたＵＩ５０において、ボタン５２～５４のいずれかを選択することで、そのボタンが示す領域を処理対象音声データから削除することができる。ここで、ボタン５２が選択されると所定閾値未満の類似度スコアに該当するデータが処理対象音声データから削除される。ボタン５３、５４に関しては事前にグラフ５８上で時間領域を選択しておくことで、選択可能な状態となる。また、この選択時には領域ｃｓ１，ｃｓ２などの時間区間を選択することも可能に構成しておくとよい。

[0072] また、声紋照合対象ファイルを最終的に生成する場合には、声紋照合対象ファイル作成ボタン５７を選択することで、処理対象音声データからそのよ

うな領域の削除がなされた結果のデータが声紋照合対象ファイルとして、保存させることができる。保存先は、サーバ１０の記憶部１２とユーザ端末２０の記憶部２２とのうち少なくとも一方とすることができるが、他の保存先であってもよく、保存先もＵＩ５０で選択可能にしておいてもよい。

[0073] 次に、本実施形態の効果に関連し、比較例として、図１２及び図１３を参照しながら、それぞれ処理対象音声データとして録音データのみ、動画データを採用した作業例について説明する。図１２は、比較例１に係る声紋照合対象ファイルの作成作業を説明するためのフロー図で、図１３は、比較例２に係る声紋照合対象ファイルの作成作業を説明するためのフロー図である。

[0074] 図１２で示す作業例は、処理対象録音データから声紋照合対象ファイルを作成する作業の手順を示している。

[0075] まず、ユーザは処理対象録音データを再生させ（ステップＳ１１）、それを聴く。また、比較例１では、処理対象録音データは、一部又は全てについてテキストへの書き起こしデータを生成したものを保存しておく。ユーザは、処理対象録音データに対象者が明確である書き起こしデータが参照できるか（存在するか）否かを判断し（ステップＳ１２）、参照できない場合にはその録音データを破棄し（ステップＳ１８）、作業を終了する。ステップＳ１２でＹＥＳの場合、ユーザは対象者が発話している区間のみを検聴し、波形を区切る作業を行う（ステップＳ１３）。

[0076] 次いで、ユーザは、ステップＳ１３の処理後の録音データに対し、対象者の声が他者の声や雑音と重なっているか否かを判断し（ステップＳ１４）、重なっている場合、録音データにおける重なっている区間を破棄する（ステップＳ１５）。これにより、ステップＳ１３の処理後の録音データが部分的に削除されることになる。次いで、対象者の声の時間長が例えば５ｓ等の規定値以上存在するか否かを判断する（ステップＳ１６）。なお、ステップＳ１４でＮＯの場合にはステップＳ１５を経ずにステップＳ１６へ進む。ステップＳ１６でＹＥＳの場合には、そこまでで残った録音データを声紋照合対象ファイルとして登録し（ステップＳ１７）、作業を終了する。一方、ステ

ップＳ１６でＮＯの場合には、規定値以上となるように、そこまで残った録音データを部品ファイルとして登録し（ステップＳ１９）、作業を終了する。

[0077] 図１３に示す比較例２は、録音データを含む動画データから声紋照合対象ファイルを作成する作業の手順を示している。

[0078] まず、ユーザは処理対象動画データを再生させ（ステップＳ２１）、それを視聴する。比較例２では映像が存在するため、ユーザは、その映像を確認し、処理対象動画データに対象者が映っていることが明確である否かを判断し（ステップＳ２２）、明確でない場合にはその動画データを処理外とし（ステップＳ２８）、作業を終了する。ステップＳ２２でＹＥＳの場合、ユーザは対象者が発話している区間のみを検聴し、波形を区切る作業を行う（ステップＳ２３）。

[0079] 次いで、ユーザは、ステップＳ２３の処理後の動画データに対し、対象者の声が他者の声や雑音と重なっているか否かを判断し（ステップＳ２４）、重なっている場合、動画データにおける重なっている区間を破棄する（ステップＳ２５）。これにより、ステップＳ２３の処理後の動画データが部分的に削除されることになる。次いで、対象者の声の時間長が例えば５ｓ等の規定値以上存在するか否かを判断する（ステップＳ２６）。なお、ステップＳ２４でＮＯの場合にはステップＳ２５を経ずにステップＳ２６へ進む。ステップＳ２６でＹＥＳの場合には、そこまでで残った動画データについての音声データを声紋照合対象ファイルとして登録し（ステップＳ２７）、作業を終了する。一方、ステップＳ２６でＮＯの場合には、そこまで残った動画データを処理外データとし（ステップＳ２８）、作業を終了する。なお、ステップＳ２６でＮＯの場合、ステップＳ１９と同様に規定値以上となるように、そこまで残った動画データを部品ファイルとして登録してから、作業を終了してもよい。

[0080] 比較例１，２のいずれの場合にも、声紋照合対象とするファイルの作成には手間と時間を要することが分かる。これに対し、本実施形態では、上述の

ような抽出処理を採用することにより、対象者が他者と会話しており且つ他者の発話が混じる区間の音声データを抽出対象外とすることができる。よって、本実施形態によれば、作業者となるユーザは提案された区間を中心に検聴することで、一部始終検聴（視聴）していた作業を、比較例１，２に比して効率化することができる。実際、例えば捜査対象の動画データや音声データには対象者以外の音声などが含まれることが多いため、本実施形態は非常に有益であると言える。

[0081] ＜変形例＞

以上、実施形態を参照して本開示を説明したが、本開示は上述の実施形態に限定されるものではない。本開示の構成や詳細には、本開示のスコープ内で当業者が理解し得る様々な変更をすることができる。そして、各実施形態は、適宜他の実施形態と組み合わせることができる。

[0082] 各図面は、１又はそれ以上の実施形態を説明するための単なる例示である。各図面は、１つの特定の実施形態のみに関連付けられるのではなく、１又はそれ以上の他の実施形態に関連付けられてもよい。当業者であれば理解できるように、いずれか１つの図面を参照して説明される様々な特徴又はステップは、例えば明示的に図示または説明されていない実施形態を作り出すために、１又はそれ以上の他の図に示された特徴又はステップと組み合わせることができる。例示的な実施形態を説明するためにいずれか１つの図に示された特徴またはステップのすべてが必ずしも必須ではなく、一部の特徴またはステップが省略されてもよい。いずれかの図に記載されたステップの順序は、適宜変更されてもよい。

[0083] また、本開示に係る声紋照合支援システムを構成する１又は複数の装置は、いずれも次のようなハードウェア構成を有することができる。図１４は、装置に含まれるハードウェア構成の一例を示す図である。

[0084] 図１４に示す装置１０００は、プロセッサ１００１、メモリ１００２、及び通信インタフェース１００３を備える。各装置の機能は、プロセッサ１００１がメモリ１００２に記憶されたプログラムを読み込んで通信インタフェ

ース１００３と協働しながら実行することにより実現されることができる。

[0085] 上述したプログラムは、コンピュータに読み込まれた場合に、実施形態で説明された１又はそれ以上の機能をコンピュータに行わせるための命令群（又はソフトウェアコード）を含む。プログラムは、非一時的なコンピュータ可読媒体又は実体のある記憶媒体に格納されてもよい。限定ではなく例として、コンピュータ可読媒体又は実体のある記憶媒体は、random-access memory (RAM)、read-only memory (ROM)、フラッシュメモリ、solid-state drive (SSD) 又はその他のメモリ技術、CD-ROM、digital versatile disk (DVD)、Blu-ray（登録商標）ディスク又はその他の光ディスクストレージ、磁気カセット、磁気テープ、磁気ディスクストレージ又はその他の磁気ストレージデバイスを含む。プログラムは、一時的なコンピュータ可読媒体又は通信媒体上で送信されてもよい。限定ではなく例として、一時的なコンピュータ可読媒体又は通信媒体は、電気的、光学的、音響的、またはその他の形式の伝搬信号を含む。

[0086] 上記の実施形態の一部又は全部は、以下の付記のようにも記載されうるが、以下には限られない。

[0087]（付記１）

声紋照合の対象者のサンプル音声データを入力する入力部と、

前記サンプル音声データに基づき、声紋照合対象音声データから前記対象者が発話している時間区間を抽出する抽出部と、

前記抽出部で抽出された前記時間区間をユーザに提示する表示部と、

を備えた声紋照合支援システム。

（付記２）

前記抽出部は、前記時間区間として、前記対象者が単独で発話している時間区間を抽出する、

付記１に記載の声紋照合支援システム。

（付記３）

前記声紋照合対象音声データから、前記抽出部で抽出された前記時間区間

以外の区間のデータを削除する削除部を備えた、

付記１又は２に記載の声紋照合支援システム。

（付記４）

前記抽出部は、前記サンプル音声データの特徴量と前記声紋照合対象音声データの特徴量との類似度スコアに基づき、前記時間区間を抽出する、

付記１～３のいずれか１項に記載の声紋照合支援システム。

（付記５）

前記抽出部は、前記時間区間として、前記声紋照合対象音声データにおける前記類似度スコアが所定閾値以上に該当する区間を抽出する、

付記４に記載の声紋照合支援システム。

（付記６）

前記所定閾値を設定する閾値設定部を備えた、

付記５に記載の声紋照合支援システム。

（付記７）

前記表示部は、前記声紋照合対象音声データに対応する時間軸あるいは前記時間軸を含む前記声紋照合対象音声データの波形を表示した状態で、少なくとも前記抽出部で抽出された前記時間区間について、前記類似度スコアを可視化して表示する、

付記４～６のいずれか１項に記載の声紋照合支援システム。

（付記８）

前記表示部は、前記声紋照合対象音声データに対応する時間軸あるいは前記時間軸を含む前記声紋照合対象音声データの波形を表示した状態で、少なくとも前記抽出部で抽出された前記時間区間について、前記類似度スコアが高い区間の疎密状態を可視化して表示する、

付記４～７のいずれか１項に記載の声紋照合支援システム。

（付記９）

前記表示部は、前記声紋照合対象音声データに対応する時間軸あるいは前記時間軸を含む前記声紋照合対象音声データの波形を表示した状態で、前記

抽出部で抽出された前記時間区間を、他の区間とは異なる表示形態で表示する、

付記１～８のいずれか１項に記載の声紋照合支援システム。

（付記１０）

前記抽出部における抽出単位となる時間間隔を設定する間隔設定部を備えた、

付記１～９のいずれか１項に記載の声紋照合支援システム。

（付記１１）

声紋照合の対象者のサンプル音声データを入力する入力処理と、

前記サンプル音声データに基づき、声紋照合対象音声データから前記対象者が発話している時間区間を抽出する抽出処理と、

抽出した前記時間区間をユーザに提示するために表示装置に表示させる表示処理と、

を備えた声紋照合支援方法。

（付記１２）

前記抽出処理は、前記時間区間として、前記対象者が単独で発話している時間区間を抽出する、

付記１１に記載の声紋照合支援方法。

（付記１３）

前記声紋照合対象音声データから、前記抽出処理で抽出された前記時間区間以外の区間のデータを削除する処理を含む、

付記１１又は１２に記載の声紋照合支援方法。

（付記１４）

前記抽出処理は、前記サンプル音声データの特徴量と前記声紋照合対象音声データの特徴量との類似度スコアに基づき実行する、

付記１１～１３のいずれか１項に記載の声紋照合支援方法。

（付記１５）

前記抽出処理は、前記時間区間として、前記声紋照合対象音声データにお

ける前記類似度スコアが所定閾値以上に該当する区間を抽出する、

付記１４に記載の声紋照合支援方法。

（付記１６）

前記所定閾値を設定する処理を含む、

付記１５に記載の声紋照合支援方法。

（付記１７）

前記表示処理は、前記表示装置において、前記声紋照合対象音声データに対応する時間軸あるいは前記時間軸を含む前記声紋照合対象音声データの波形を表示させた状態で、少なくとも前記抽出処理で抽出された前記時間区間について、前記類似度スコアを可視化して表示させる処理を含む、

付記１４～１６のいずれか１項に記載の声紋照合支援方法。

（付記１８）

前記表示処理は、前記表示装置において、前記声紋照合対象音声データに対応する時間軸あるいは前記時間軸を含む前記声紋照合対象音声データの波形を表示させた状態で、少なくとも前記抽出処理で抽出された前記時間区間について、前記類似度スコアが高い区間の疎密状態を可視化して表示させる処理を含む、

付記１４～１７のいずれか１項に記載の声紋照合支援方法。

（付記１９）

前記表示処理は、前記表示装置において、前記声紋照合対象音声データに対応する時間軸あるいは前記時間軸を含む前記声紋照合対象音声データの波形を表示させた状態で、前記抽出処理で抽出された前記時間区間を、他の区間とは異なる表示形態で表示させる処理を含む、

付記１１～１８のいずれか１項に記載の声紋照合支援方法。

（付記２０）

前記抽出処理における抽出単位となる時間間隔を設定する処理を含む、

付記１１～１９のいずれか１項に記載の声紋照合支援方法。

（付記２１）

コンピュータに、

声紋照合の対象者のサンプル音声データを入力する入力処理と、

前記サンプル音声データに基づき、声紋照合対象音声データから前記対象者が発話している時間区間を抽出する抽出処理と、

抽出した前記時間区間をユーザに提示するために表示装置に表示させる表示処理と、

を含む声紋照合支援処理を実行させるプログラム。

（付記２２）

前記抽出処理は、前記時間区間として、前記対象者が単独で発話している時間区間を抽出する、

付記２１に記載のプログラム。

（付記２３）

前記声紋照合支援処理は、前記声紋照合対象音声データから、前記抽出処理で抽出された前記時間区間以外の区間のデータを削除する処理を含む、

付記２１又は２２に記載のプログラム。

（付記２４）

前記抽出処理は、前記サンプル音声データの特徴量と前記声紋照合対象音声データの特徴量との類似度スコアに基づき実行する、

付記２１～２３のいずれか１項に記載のプログラム。

（付記２５）

前記抽出処理は、前記時間区間として、前記声紋照合対象音声データにおける前記類似度スコアが所定閾値以上に該当する区間を抽出する、

付記２４に記載のプログラム。

（付記２６）

前記声紋照合支援処理は、前記所定閾値を設定する処理を含む、

付記２５に記載のプログラム。

（付記２７）

前記表示処理は、前記表示装置において、前記声紋照合対象音声データに

対応する時間軸あるいは前記時間軸を含む前記声紋照合対象音声データの波形を表示させた状態で、少なくとも前記抽出処理で抽出された前記時間区間について、前記類似度スコアを可視化して表示させる処理を含む、

付記２４～２６のいずれか１項に記載のプログラム。

（付記２８）

前記表示処理は、前記表示装置において、前記声紋照合対象音声データに対応する時間軸あるいは前記時間軸を含む前記声紋照合対象音声データの波形を表示させた状態で、少なくとも前記抽出処理で抽出された前記時間区間について、前記類似度スコアが高い区間の疎密状態を可視化して表示させる処理を含む、

付記２４～２７のいずれか１項に記載のプログラム。

（付記２９）

前記表示処理は、前記表示装置において、前記声紋照合対象音声データに対応する時間軸あるいは前記時間軸を含む前記声紋照合対象音声データの波形を表示させた状態で、前記抽出処理で抽出された前記時間区間を、他の区間とは異なる表示形態で表示させる処理を含む、

付記２１～２８のいずれか１項に記載のプログラム。

（付記３０）

前記声紋照合支援処理は、前記抽出処理における抽出単位となる時間間隔を設定する処理を含む、

付記２１～２９のいずれか１項に記載のプログラム。

[0088] この出願は、２０２２年１２月５日に出願された日本出願特願2022-194509を基礎とする優先権を主張し、その開示の全てをここに取り込む。

符号の説明

[0089] １、１００　声紋照合支援システム

１ａ　入力部

１ｂ　抽出部

１ｃ　表示部

１０　サーバ

１１　制御部

１２　記憶部

１３　通信部

２０　ユーザ端末

２１　制御部

２２　記憶部

２３　通信部

２４　表示部

１０００　装置

１００１　プロセッサ

１００２　メモリ

１００３　通信インタフェース

請求の範囲

[請求項1] 声紋照合の対象者のサンプル音声データを入力する入力手段と、

前記サンプル音声データに基づき、声紋照合対象音声データから前記対象者が発話している時間区間を抽出する抽出手段と、

前記抽出手段で抽出された前記時間区間をユーザに提示する表示手段と、

を備えた声紋照合支援システム。

[請求項2] 前記抽出手段は、前記時間区間として、前記対象者が単独で発話している時間区間を抽出する、

請求項１に記載の声紋照合支援システム。

[請求項3] 前記声紋照合対象音声データから、前記抽出手段で抽出された前記時間区間以外の区間のデータを削除する削除手段を備えた、

請求項１又は２に記載の声紋照合支援システム。

[請求項4] 前記抽出手段は、前記サンプル音声データの特徴量と前記声紋照合対象音声データの特徴量との類似度スコアに基づき、前記時間区間を抽出する、

請求項１又は２に記載の声紋照合支援システム。

[請求項5] 前記抽出手段は、前記時間区間として、前記声紋照合対象音声データにおける前記類似度スコアが所定閾値以上に該当する区間を抽出する、

請求項４に記載の声紋照合支援システム。

[請求項6] 前記所定閾値を設定する閾値設定手段を備えた、

請求項５に記載の声紋照合支援システム。

[請求項7] 前記表示手段は、前記声紋照合対象音声データに対応する時間軸あるいは前記時間軸を含む前記声紋照合対象音声データの波形を表示した状態で、前記抽出手段で抽出された前記時間区間を、他の区間とは異なる表示形態で表示する、

請求項１又は２に記載の声紋照合支援システム。

[請求項8] 前記抽出手段における抽出単位となる時間間隔を設定する間隔設定手段を備えた、

請求項１又は２に記載の声紋照合支援システム。

[請求項9] 前記表示手段は、前記声紋照合対象音声データに対応する時間軸あるいは前記時間軸を含む前記声紋照合対象音声データの波形を表示した状態で、前記抽出手段で抽出された前記時間区間を、他の区間とは異なる表示形態で表示する、

請求項１又は２に記載の声紋照合支援システム。

[請求項10] 前記抽出手段における抽出単位となる時間間隔を設定する間隔設定手段を備えた、

請求項１又は２に記載の声紋照合支援システム。

[請求項11] 声紋照合の対象者のサンプル音声データを入力する入力処理と、

前記サンプル音声データに基づき、声紋照合対象音声データから前記対象者が発話している時間区間を抽出する抽出処理と、

抽出した前記時間区間をユーザに提示するために表示装置に表示させる表示処理と、

を備えた声紋照合支援方法。

[請求項12] 前記抽出処理は、前記時間区間として、前記対象者が単独で発話している時間区間を抽出する、

請求項１１に記載の声紋照合支援方法。

[請求項13] 前記声紋照合対象音声データから、前記抽出処理で抽出された前記時間区間以外の区間のデータを削除する処理を含む、

請求項１１又は１２に記載の声紋照合支援方法。

[請求項14] 前記抽出処理は、前記サンプル音声データの特徴量と前記声紋照合対象音声データの特徴量との類似度スコアに基づき実行する、

請求項１１又は１２に記載の声紋照合支援方法。

[請求項15] 前記抽出処理は、前記時間区間として、前記声紋照合対象音声データにおける前記類似度スコアが所定閾値以上に該当する区間を抽出す

る、

請求項１４に記載の声紋照合支援方法。

[請求項16] コンピュータに、

声紋照合の対象者のサンプル音声データを入力する入力処理と、

前記サンプル音声データに基づき、声紋照合対象音声データから前記対象者が発話している時間区間を抽出する抽出処理と、

抽出した前記時間区間をユーザに提示するために表示装置に表示させる表示処理と、

を含む声紋照合支援処理を実行させるプログラム。

[請求項17] 前記抽出処理は、前記時間区間として、前記対象者が単独で発話している時間区間を抽出する、

請求項１６に記載のプログラム。

[請求項18] 前記声紋照合支援処理は、前記声紋照合対象音声データから、前記抽出処理で抽出された前記時間区間以外の区間のデータを削除する処理を含む、

請求項１６又は１７に記載のプログラム。

[請求項19] 前記抽出処理は、前記サンプル音声データの特徴量と前記声紋照合対象音声データの特徴量との類似度スコアに基づき実行する、

請求項１６又は１７に記載のプログラム。

[請求項20] 前記抽出処理は、前記時間区間として、前記声紋照合対象音声データにおける前記類似度スコアが所定閾値以上に該当する区間を抽出する、

請求項１９に記載のプログラム。

[図1]

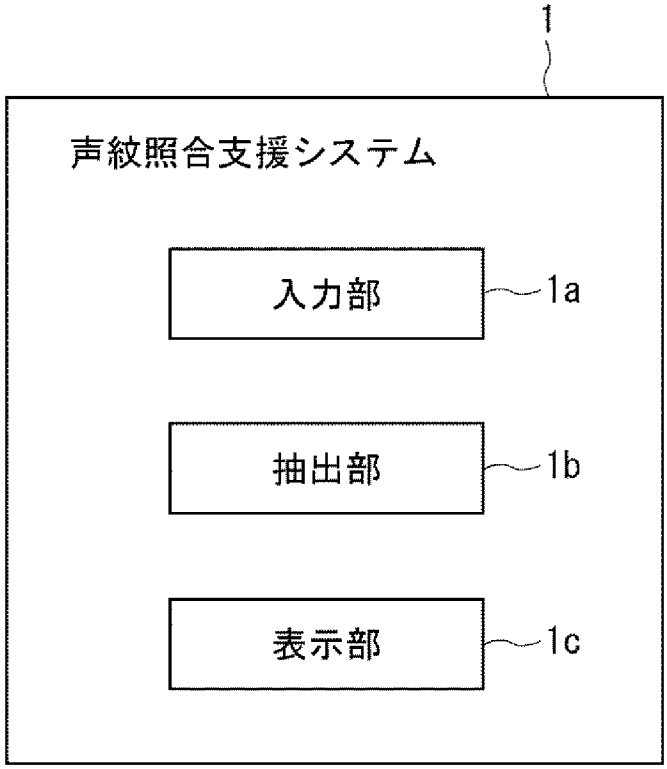


Fig. 1

[図2]

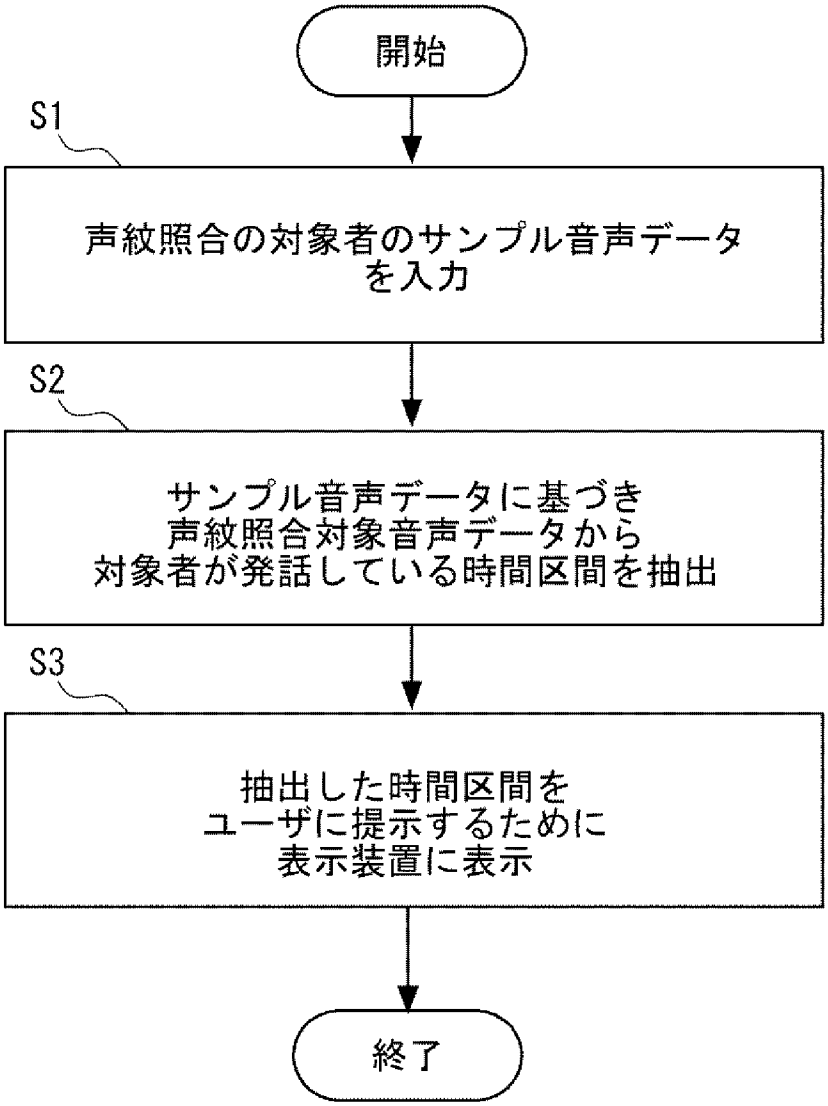

開始
S1
声紋照合の対象者のサンプル音声データ
を入力
S2
サンプル音声データに基づき
声紋照合対象音声データから
対象者が発話している時間区間を抽出
S3
抽出した時間区間を
ユーザに提示するために
表示装置に表示
終了

Fig. 2

[図3]

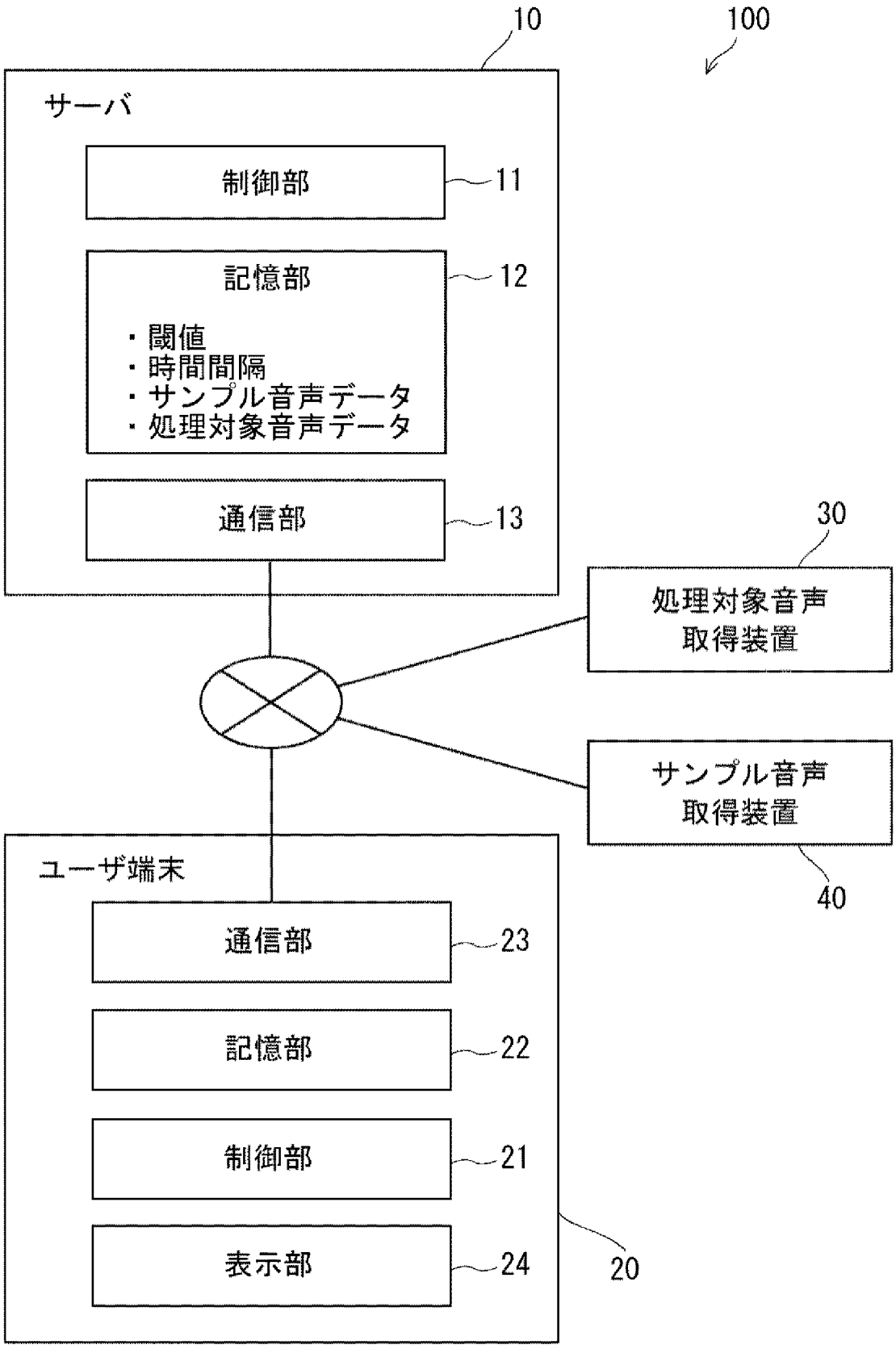

10
100
サーバ
制御部
11
記憶部
12
・閾値
・時間間隔
・サンプル音声データ
・処理対象音声データ
通信部
13
30
処理対象音声
取得装置
サンプル音声
取得装置
40
ユーザ端末
通信部
23
記憶部
22
制御部
21
表示部
24
20

Fig. 3

[図4]

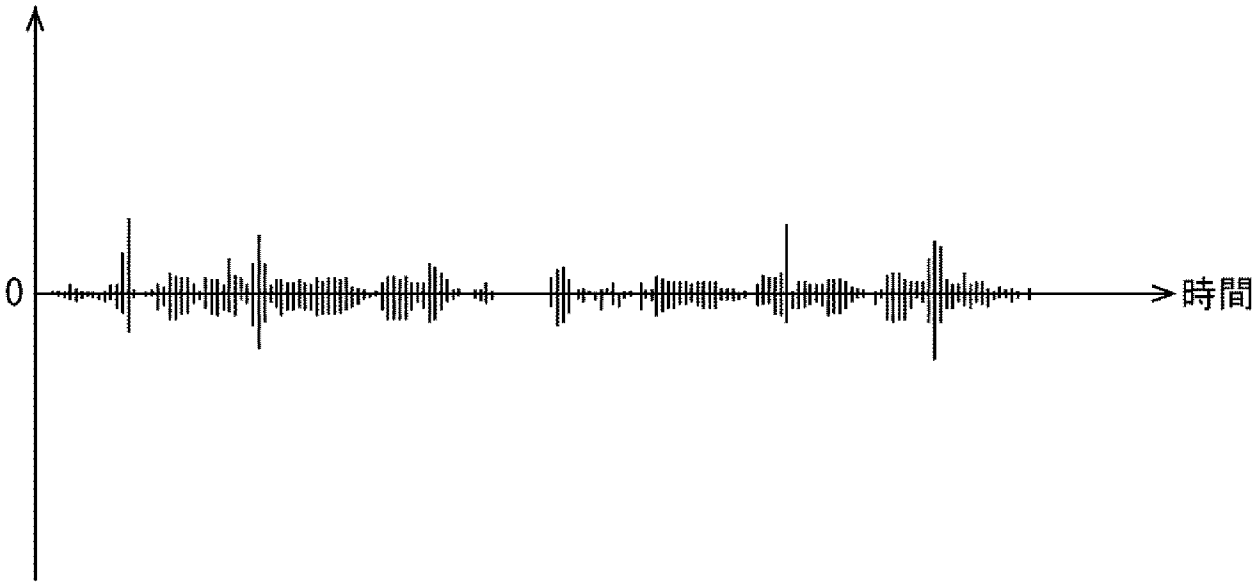


Fig. 4

[図5]

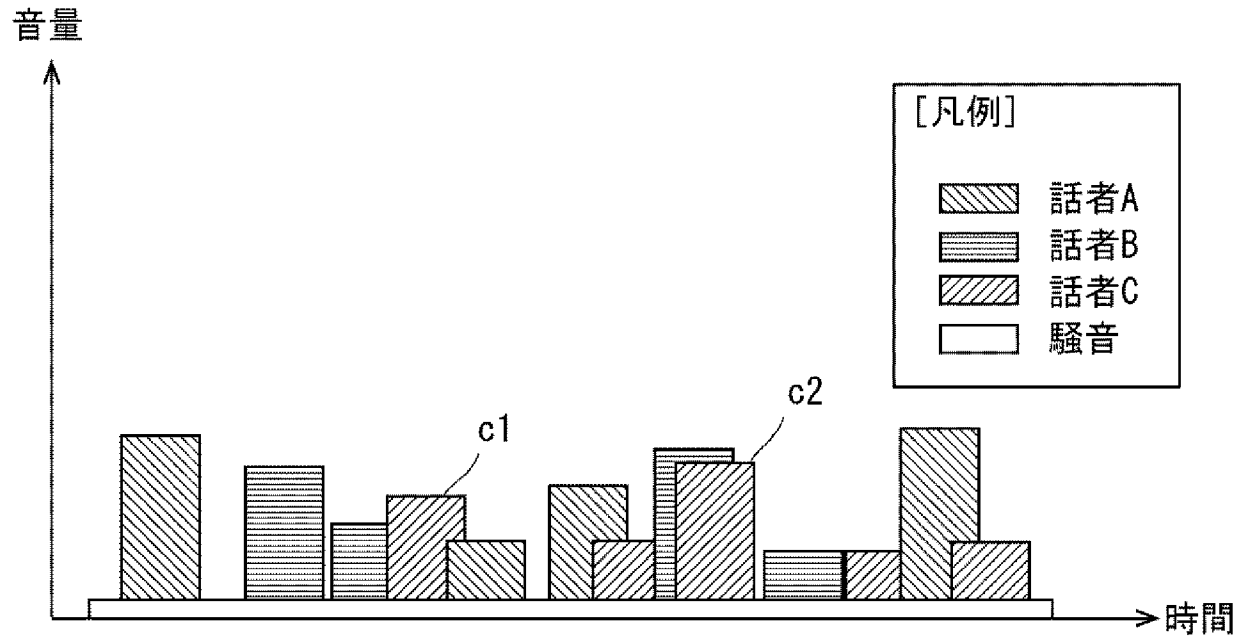


Fig. 5

[図6]

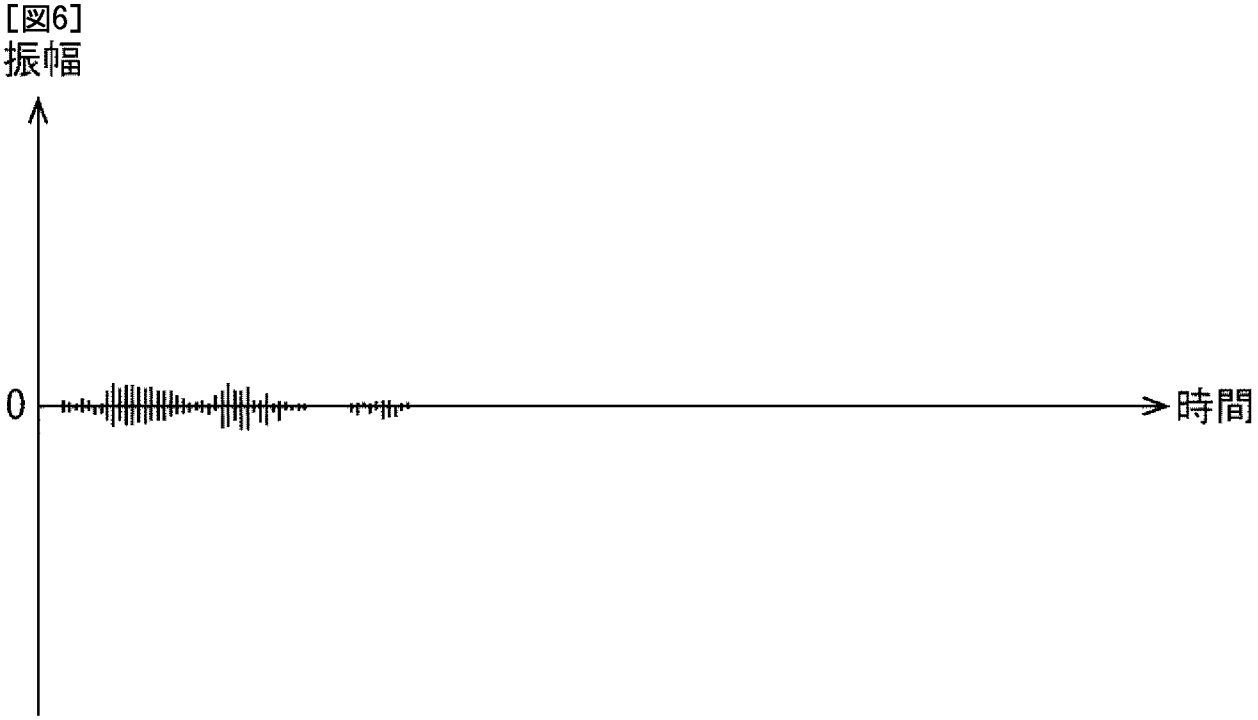


Fig. 6

[図7]

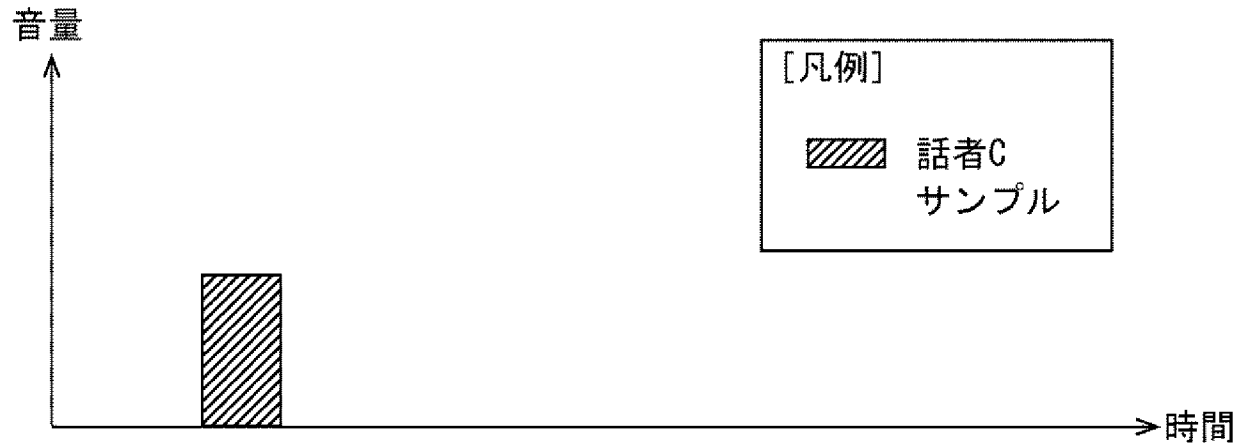


Fig. 7

[図8]

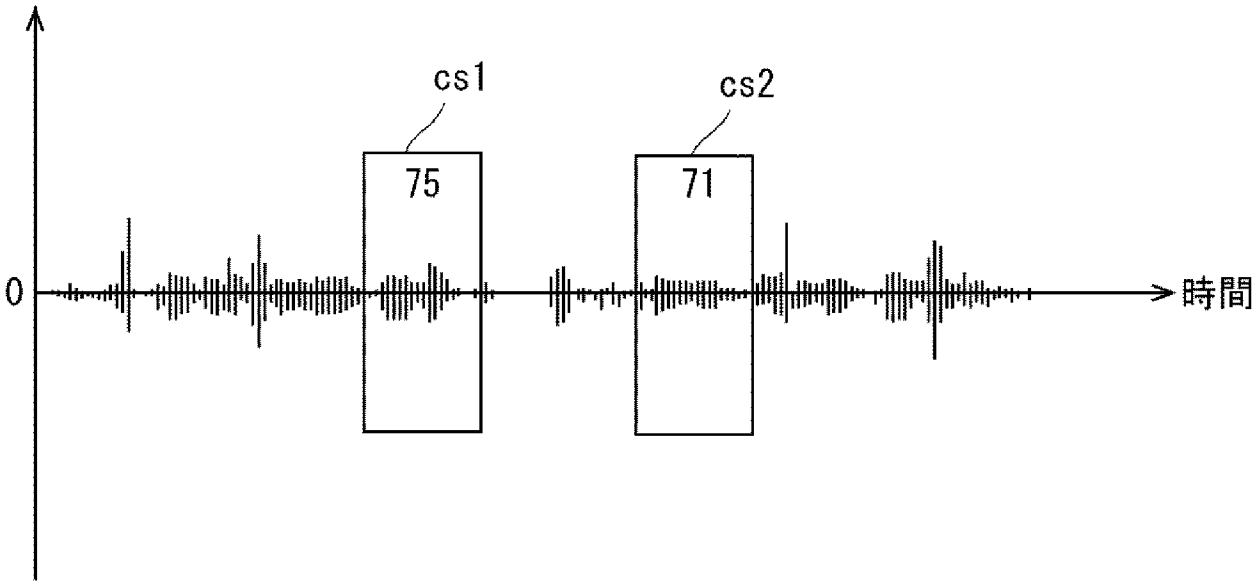


Fig. 8

[図9]

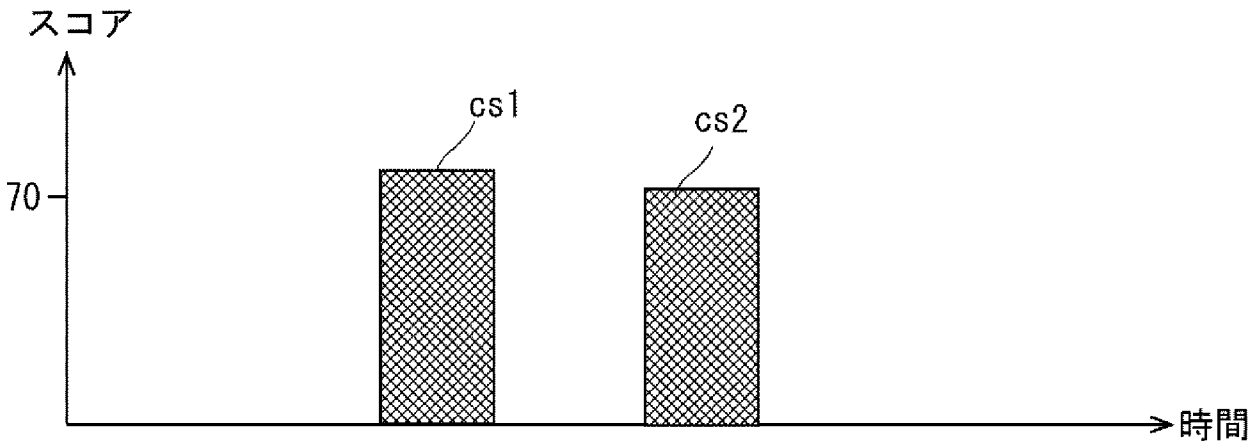


Fig. 9

[図10]

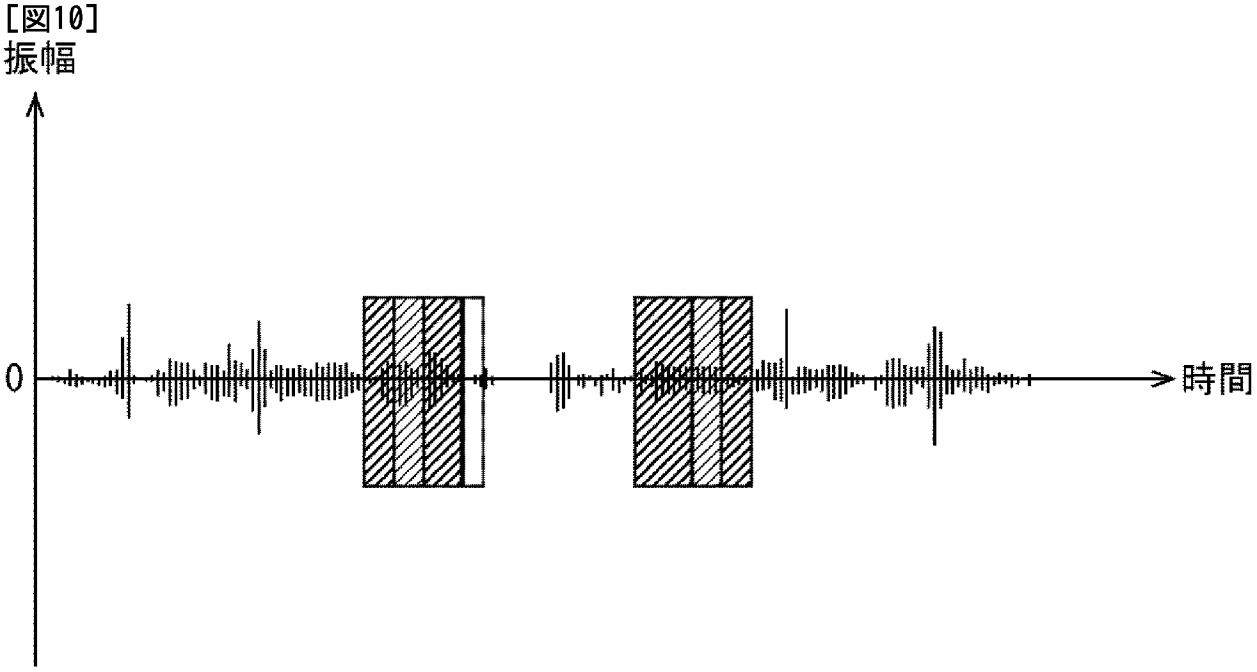


Fig. 10

[図11]

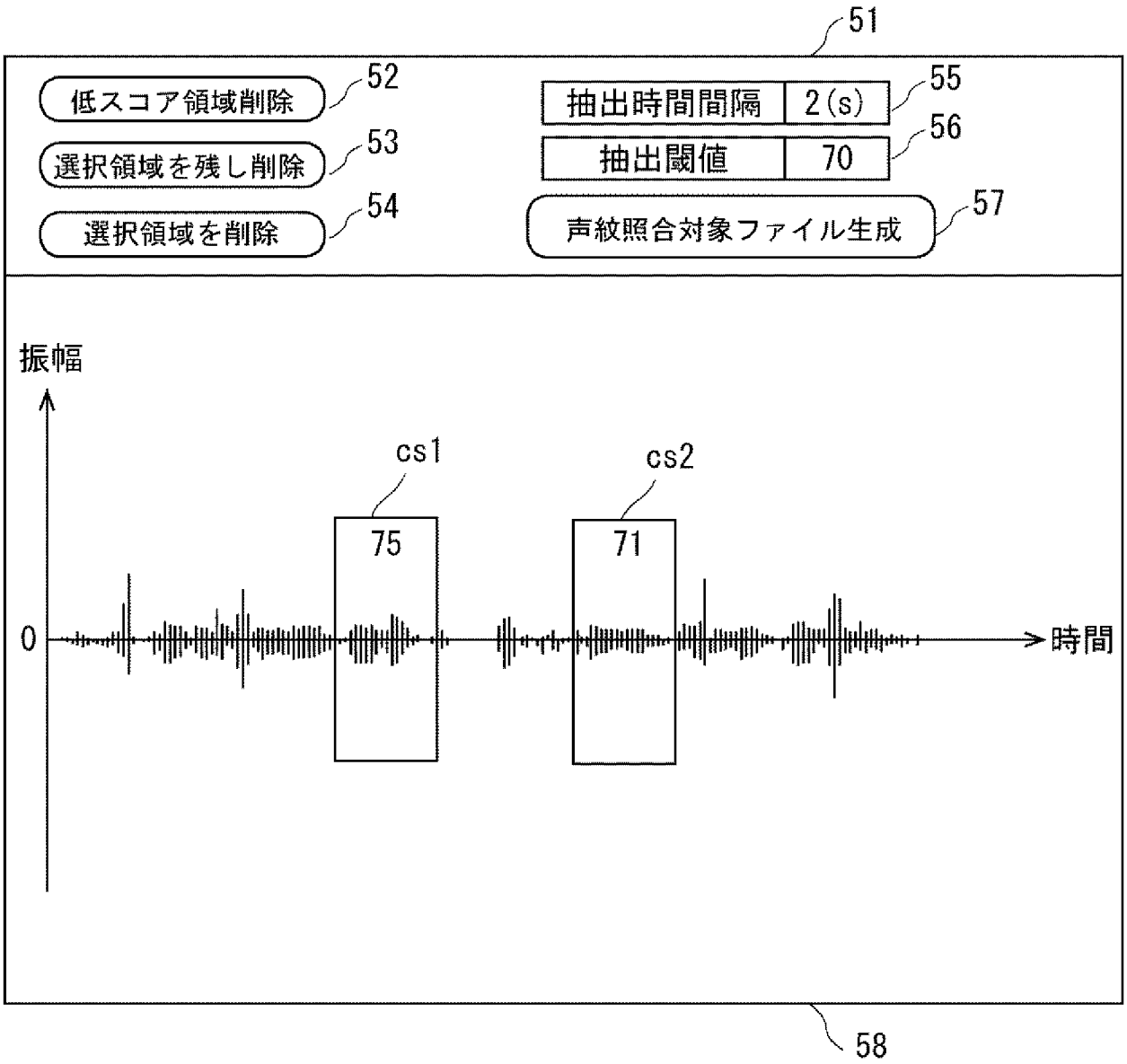

50
51
低スコア領域削除
52
選択領域を残し削除
53
選択領域を削除
54
抽出時間間隔
2(s)
55
抽出閾値
70
56
声紋照合対象ファイル生成
57
振幅
cs1
cs2
75
71
0
時間
58

Fig. 11

[図12]

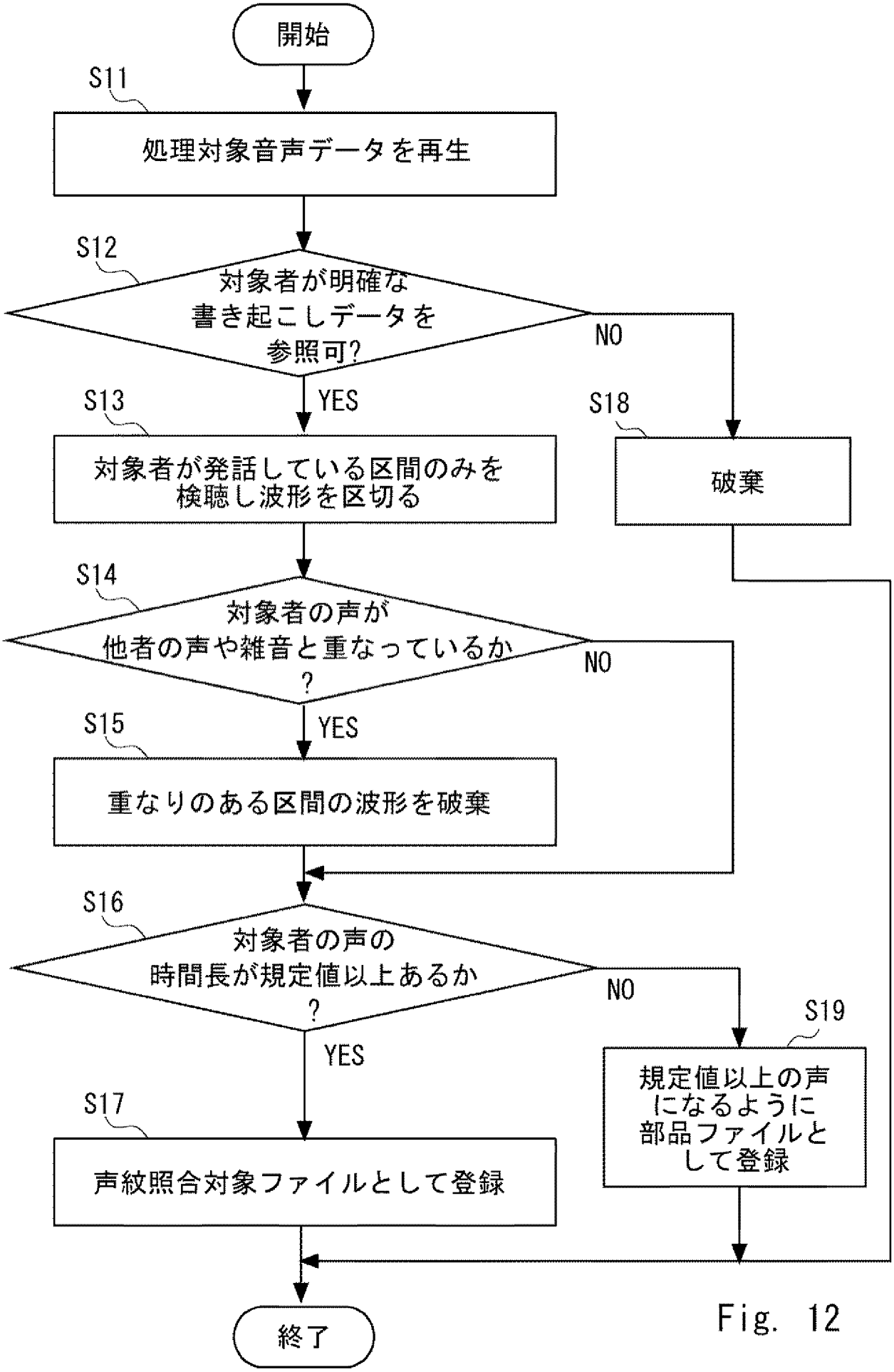

開始
S11
処理対象音声データを再生
S12
対象者が明確な書き起こしデータを参照可?
NO
YES
S13
対象者が発話している区間のみを検聴し波形を区切る
S18
破棄
S14
対象者の声が他者の声や雑音と重なっているか?
NO
YES
S15
重なりのある区間の波形を破棄
S16
対象者の声の時間長が規定値以上あるか?
NO
YES
S19
規定値以上の声になるように部品ファイルとして登録
S17
声紋照合対象ファイルとして登録
終了

Fig. 12

[図13]

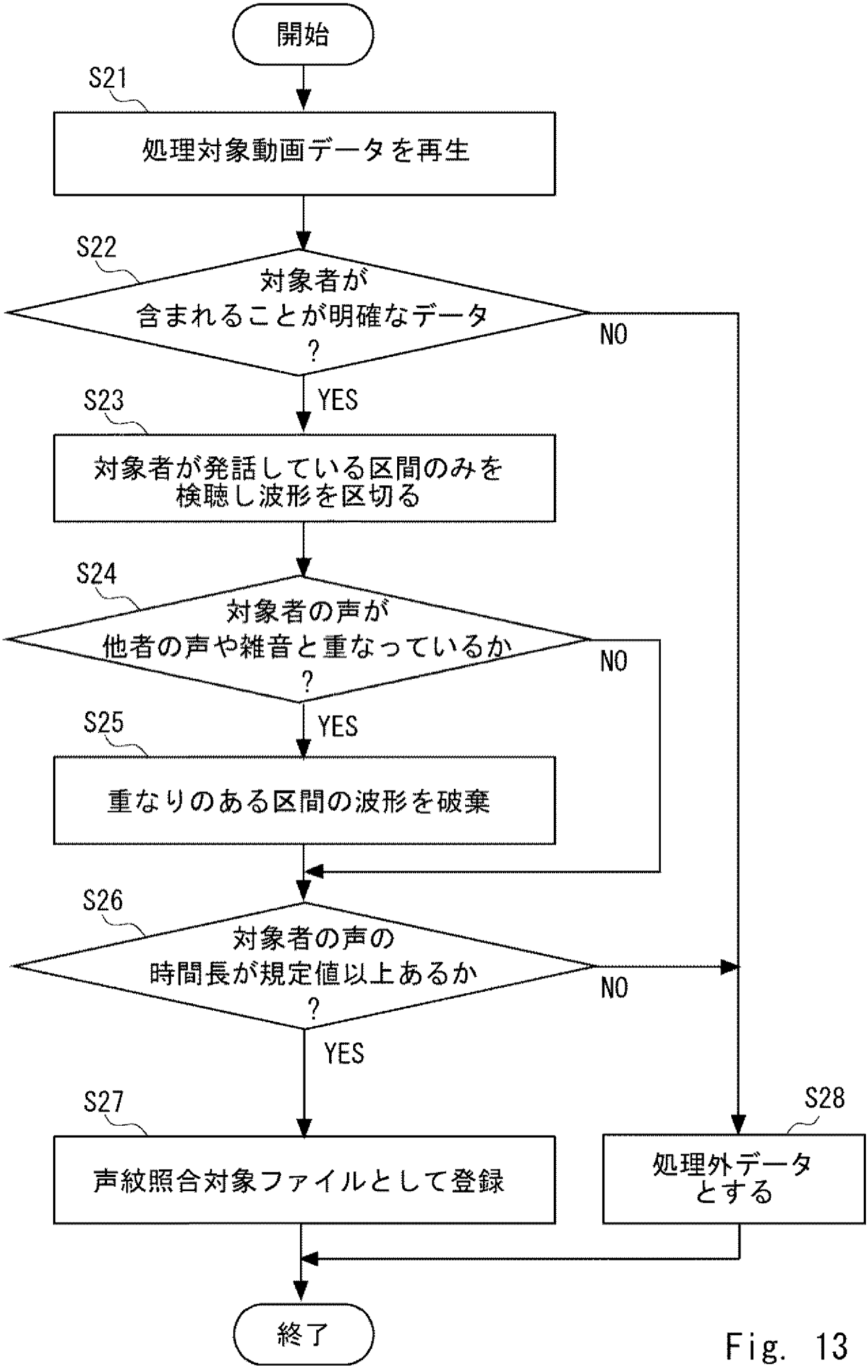

開始
S21
処理対象動画データを再生
S22
対象者が
含まれることが明確なデータ
?
NO
YES
S23
対象者が発話している区間のみを
検聴し波形を区切る
S24
対象者の声が
他者の声や雑音と重なっているか
?
NO
YES
S25
重なりのある区間の波形を破棄
S26
対象者の声の
時間長が規定値以上あるか
?
NO
YES
S27
声紋照合対象ファイルとして登録
S28
処理外データ
とする
終了

Fig. 13

[図14]

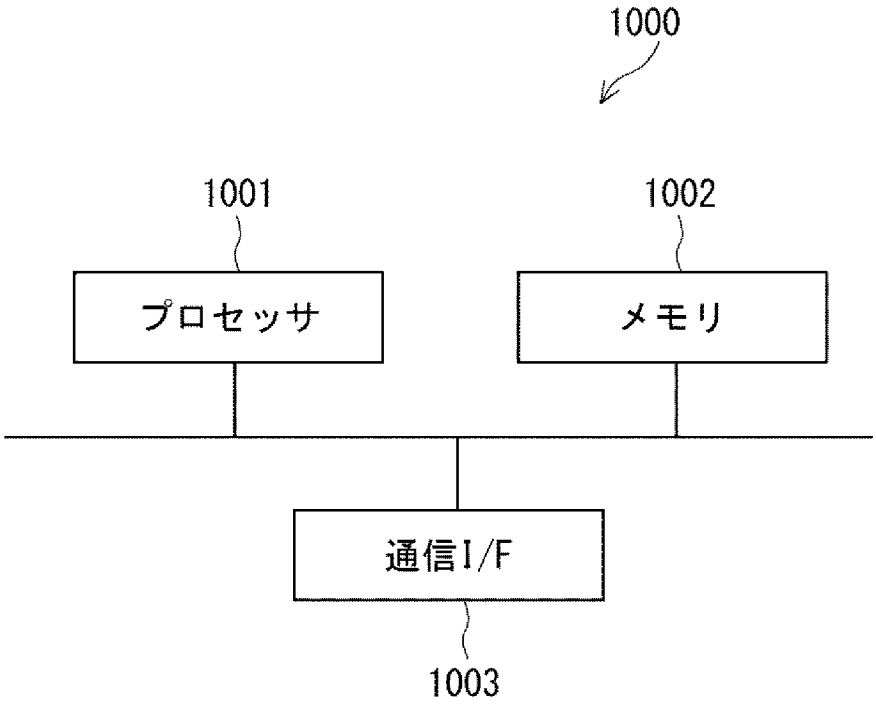


Fig. 14

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2023/041303

A. CLASSIFICATION OF SUBJECT MATTER

G10L 17/00(2013.01)i; ***G10L 15/04***(2013.01)i; ***G10L 17/08***(2013.01)i
FI: G10L17/00 200C; G10L17/08; G10L15/04 300C

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)

G10L15/00-17/26

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 1971-2024
Registered utility model specifications of Japan 1996-2024
Published registered utility model applications of Japan 1994-2024

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.
X	JP 2008-32825 A (FUJITSU FSAS INC.) 14 February 2008 (2008-02-14) paragraphs [0010]-[0060]	1-2, 11-12, 16-17
Y		3-10, 13-15, 18-20
Y	JP 2005-321530 A (SONY CORPORATION) 17 November 2005 (2005-11-17) paragraphs [0011]-[0028]	3, 7-10, 13, 18
Y	JP 2005-326576 A (SONY CORPORATION) 24 November 2005 (2005-11-24) paragraphs [0015]-[0041]	4-6, 14-15, 19-20
A	JP 2010-113167 A (NEIKUSU K.K.) 20 May 2010 (2010-05-20) entire text	1-20
A	WO 2020/003413 A1 (NEC CORPORATION) 02 January 2020 (2020-01-02) entire text	1-20
A	US 2021/0326421 A1 (PINDROP SECURITY, INC.) 21 October 2021 (2021-10-21) entire text	1-20

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

Date of the actual completion of the international search	Date of mailing of the international search report
25 January 2024	**06 February 2024**

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

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/JP2023/041303

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
JP 2008-32825 A	14 February 2008	(Family: none)	
JP 2005-321530 A	17 November 2005	(Family: none)	
JP 2005-326576 A	24 November 2005	(Family: none)	
JP 2010-113167 A	20 May 2010	(Family: none)	
WO 2020/003413 A1	02 January 2020	US 2021/0287682 A1 entire text EP 3816996 A1	
US 2021/0326421 A1	21 October 2021	WO 2021/211836 A1 EP 4136637 A1 KR 10-2023-0011307 A JP 2023-521768 A	

国際調査報告

国際出願番号
PCT/JP2023/041303

A. 発明の属する分野の分類（国際特許分類（ＩＰＣ））

G10L 17/00(2013.01)i; G10L 15/04(2013.01)i; G10L 17/08(2013.01)i
FI: G10L17/00 200C; G10L17/08; G10L15/04 300C

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））

G10L15/00-17/26

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	１９７１－２０２４年
日本国実用新案登録公報	１９９６－２０２４年
日本国登録実用新案公報	１９９４－２０２４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
X	JP 2008-32825 A (株式会社富士通エフサス) 14.02.2008 (2008-02-14) 段落[0010]-[0060]	1-2,11-12,16-17
Y		3-10,13-15,18-20
Y	JP 2005-321530 A (ソニー株式会社) 17.11.2005 (2005-11-17) 段落[0011]-[0028]	3,7-10,13,18
Y	JP 2005-326576 A (ソニー株式会社) 24.11.2005 (2005-11-24) 段落[0015]-[0041]	4-6,14-15,19-20
A	JP 2010-113167 A (株式会社ネイクス) 20.05.2010 (2010-05-20) 全文	1-20
A	WO 2020/003413 A1 (日本電気株式会社) 02.01.2020 (2020-01-02) 全文	1-20
A	US 2021/0326421 A1 (PINDROP SECURITY, INC.) 21.10.2021 (2021-10-21) 全文	1-20

☐ 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
"D" 国際出願で出願人が先行技術文献として記載した文献
"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抵触するものではなく、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	国際調査報告の発送日
25.01.2024	06.02.2024

名称及びあて先

日本国特許庁(ISA/JP)
〒100-8915
日本国
東京都千代田区霞が関三丁目４番３号

権限のある職員（特許庁審査官）

山下　剛史 5Z 8946

電話番号 03-3581-1101 内線 3591

国際調査報告
パテントファミリーに関する情報

国際出願番号
PCT/JP2023/041303

引用文献	公表日	パテントファミリー文献	公表日
JP 2008-32825 A	14.02.2008	（ファミリーなし）	
JP 2005-321530 A	17.11.2005	（ファミリーなし）	
JP 2005-326576 A	24.11.2005	（ファミリーなし）	
JP 2010-113167 A	20.05.2010	（ファミリーなし）	
WO 2020/003413 A1	02.01.2020	US 2021/0287682 A1 全文 EP 3816996 A1	
US 2021/0326421 A1	21.10.2021	WO 2021/211836 A1 EP 4136637 A1 KR 10-2023-0011307 A JP 2023-521768 A	