(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第6641832号
(P6641832)
(45)発行日　令和2年2月5日(2020.2.5)　(24)登録日　令和2年1月8日(2020.1.8)

(51)Int.Cl.　　F I
G10L 25/51 (2013.01)　G10L 25/51
G10L 25/06 (2013.01)　G10L 25/06
G10L 15/10 (2006.01)　G10L 15/10　500Z

請求項の数 13　(全 31 頁)

(21)出願番号　特願2015-186617(P2015-186617)
(22)出願日　平成27年9月24日(2015.9.24)
(65)公開番号　特開2017-62307(P2017-62307A)
(43)公開日　平成29年3月30日(2017.3.30)
審査請求日　平成30年6月8日(2018.6.8)

(73)特許権者　000005223
富士通株式会社
神奈川県川崎市中原区上小田中４丁目１番１号
(74)代理人　100147164
弁理士　向山　直樹
(72)発明者　外川　太郎
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
(72)発明者　香村　紗友梨
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
(72)発明者　大谷　猛
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
最終頁に続く

(54)【発明の名称】音声処理装置、音声処理方法および音声処理プログラム

(57)【特許請求の範囲】
【請求項１】
第１音声が含まれる第１入力信号と、第２音声が含まれる第２入力信号を取得する取得部と、
前記第１入力信号の第１信号強度と、前記第２入力信号の第２信号強度を検出する検出部と、
前記第１信号強度の時系列と前記第２信号強度の時系列との相関係数を算出する算出部と、
前記相関係数に基づいて、前記第１音声と前記第２音声が会話している状態か否かを判定する判定部と、を有し、
前記検出部は、
前記第１入力信号に含まれる第１発話区間を前記第１信号強度に基づいて検出し、
前記第２入力信号に含まれる第２発話区間を前記第２信号強度に基づいて検出し、
前記第１発話区間と前記第２発話区間が重複する重複発話区間を検出し、
前記算出部は、
前記重複発話区間が所定の第２閾値未満の前記第１信号強度と前記第２信号強度以外の前記第１信号強度の時系列と前記第２信号強度の時系列との前記相関係数を算出することを特徴とする音声処理装置。
【請求項２】
前記判定部は、前記相関係数が負であり、前記相関係数が所定の第１閾値未満の場合に

、前記第１音声と前記第２音声が会話している状態と判定することを特徴とする請求項１記載の音声処理装置。
【請求項３】
前記第１信号強度は前記第１音声のパワーまたは信号対雑音比であり、前記第２信号強度は前記第２音声のパワーまたは信号対雑音比であることを特徴とする請求項１または請求項２記載の音声処理装置。
【請求項４】
前記検出部は、
前記第１入力信号に含まれる第１無音区間を前記第１信号強度に基づいて検出し、
前記第２入力信号に含まれる第２無音区間を前記第２信号強度に基づいて検出し、
前記第１無音区間と前記第２無音区間が重複する重複無音区間を検出し、
前記算出部は、
前記重複無音区間が所定の第３閾値未満の前記第１信号強度と前記第２信号強度を、前記相関係数の算出に用いないことを特徴とする請求項１ないし請求項３の何れか一項に記載の音声処理装置。
【請求項５】
前記算出部は、前記第１信号強度の第１位相または、前記第２信号強度の第２位相を所定の範囲で変化させて複数の前記相関係数を算出し、
前記判定部は、複数の前記相関係数のうち最小値となる前記相関係数に基づいて、前記第１音声と前記第２音声が会話している状態か否かを判定することを特徴とする請求項１ないし請求項４の何れか一項に記載の音声処理装置。
【請求項６】
前記検出部は、
前記第１信号強度の時系列と前記第２信号強度の時系列との大小関係を比較し、
前記大小関係が反転する反転回数が所定の第４閾値以上となる相関係数算出区間を検出し、
前記算出部は、
前記相関係数算出区間における前記第１信号強度の時系列と前記第２信号強度の時系列との前記相関係数を算出することを特徴とする請求項１ないし請求項５の何れか一項に記載の音声処理装置。
【請求項７】
前記取得部は、第３音声が含まれる第３入力信号を更に取得し、
前記検出部は、前記第３入力信号の第３信号強度を更に検出し、
前記算出部は、前記第１信号強度の時系列、前記第２信号強度の時系列または前記第３信号強度の時系列のうち、２つの信号強度の組み合わせに対する複数の前記相関係数を算出し、
前記判定部は、複数の前記相関係数のうち、前記相関係数が最小値となる２つの前記信号強度の組み合わせに基づいて、前記第１音声、前記第２音声または前記第３音声から、会話している音声の組み合わせを判定することを特徴とする請求項１ないし請求項６の何れか一項に記載の音声処理装置。
【請求項８】
前記算出部は、前記最小値となる前記相関係数の算出に用いた２つの前記信号強度を加算した加算信号強度を算出し、
前記加算信号強度の時系列と、前記最小値となる前記相関係数の算出に用いた２つの前記信号強度以外の１つの信号強度の時系列との参照相関係数を算出し、
前記判定部は、前記相関係数が負であり、前記相関係数が所定の第１閾値未満の場合に、前記第１音声と前記第２音声が会話している状態と判定し、
前記最小値となる前記相関係数が前記参照相関係数を下回る場合、または、前記第１閾値未満の場合に、前記参照相関係数の算出に用いた３つの前記信号強度に基づいて、会話している音声の組み合わせを判定することを特徴とする請求項７記載の音声処理装置。

【請求項９】
前記検出部は、
前記第１入力信号に含まれる第１無音区間を前記第１信号強度に基づいて検出し、
前記第２入力信号に含まれる第２無音区間を前記第２信号強度に基づいて検出し、
前記第１無音区間と前記第２無音区間が重複する重複無音区間を検出し、
前記算出部は、前記最小値となる前記相関係数の算出に用いた２つの前記信号強度の前記重複無音区間が第５閾値以上の場合に、前記参照相関係数を算出することを特徴とする請求項８記載の音声処理装置。
【請求項１０】
前記検出部は、
前記第１入力信号に含まれる前記第２信号強度または、前記第２入力信号に含まれる前記第１信号強度を更に検出し、
前記算出部は、
前記第１入力信号に含まれる前記第２信号強度の時系列と前記第２入力信号に含まれる前記第２信号強度の時系列との第２相関係数、または、
前記第２入力信号に含まれる前記第１信号強度の時系列と前記第１入力信号に含まれる前記第１信号強度の時系列との第３相関係数を更に算出し、
前記判定部は、前記第２相関係数または前記第３相関係数に基づいて、前記第１音声と前記第２音声が会話している状態か否かを判定することを特徴とする請求項１ないし請求項９の何れか一項に記載の音声処理装置。
【請求項１１】
前記判定部は、前記相関係数が正であり、前記第２相関係数が所定の第６閾値以上の場合に、前記第１音声と前記第２音声が前記会話している状態と判定することを特徴とする請求項１０記載の音声処理装置。
【請求項１２】
第１音声が含まれる第１入力信号と、第２音声が含まれる第２入力信号を取得し、
前記第１入力信号の第１信号強度と、前記第２入力信号の第２信号強度を検出し、
前記第１入力信号に含まれる第１発話区間を前記第１信号強度に基づいて検出し、
前記第２入力信号に含まれる第２発話区間を前記第２信号強度に基づいて検出し、
前記第１発話区間と前記第２発話区間が重複する重複発話区間を検出し、
前記重複発話区間が所定の第２閾値未満の前記第１信号強度と前記第２信号強度以外の前記第１信号強度の時系列と前記第２信号強度の時系列との相関係数を算出し、
前記相関係数に基づいて、前記第１音声と前記第２音声が会話している状態か否かを判定することを含む音声処理方法。
【請求項１３】
コンピュータに
第１音声が含まれる第１入力信号と、第２音声が含まれる第２入力信号を取得し、
前記第１入力信号の第１信号強度と、前記第２入力信号の第２信号強度を検出し、
前記第１入力信号に含まれる第１発話区間を前記第１信号強度に基づいて検出し、
前記第２入力信号に含まれる第２発話区間を前記第２信号強度に基づいて検出し、
前記第１発話区間と前記第２発話区間が重複する重複発話区間を検出し、
前記重複発話区間が所定の第２閾値未満の前記第１信号強度と前記第２信号強度以外の前記第１信号強度の時系列と前記第２信号強度の時系列との相関係数を算出し、
前記相関係数に基づいて、前記第１音声と前記第２音声が会話している状態か否かを判定する
ことを実行させることを特徴とする音声処理プログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、例えば、音声処理装置、音声処理方法および音声処理プログラムに関する。

【背景技術】
【０００２】
近年、例えば、企業内の従業員同士が円滑なコミュニケーションを実現できているか否かを把握する為に、録音機器（例えばマイクロフォン）を従業員毎に装着して、各従業員の音声を常時録音することが行われている。従業員の音声の常時録音によって、会議や雑談などの対面での会話音声（音声データと称しても良い）や、電話などの通信を介した会話音声など、様々なコミュニケーションにおける会話音声を取得することが出来る。常時録音された音声データから誰と誰がどれ位の時間に渡って会話したのかを分析することで、企業内の従業員同士が円滑なコミュニケーションを実現できているか否かを把握することが可能となる。
【先行技術文献】
【特許文献】
【０００３】
【特許文献１】特開２０１０－２６６５２２号公報
【発明の概要】
【発明が解決しようとする課題】
【０００４】
常時録音された会話音声から誰と誰がどれ位の時間会話したのかを分析する場合、複数の話者の音声が個別に録音された音声データから、実際に会話が行われている話者に対応した音声データの組み合せを判定する必要がある。この場合、評価者による主観評価により手作業で音声データの組み合せの判定作業を実施する必要がある。また、組み合せの判定作業を行う音声データが多い場合、音声データに対応する話者を特定することが評価者にとって困難な場合も想定され得る。一方、複数の話者の音声が個別に録音された音声データから、会話が行われている話者に対応した音声データの組み合せを判定することが可能となる音声処理装置は、実現されていない状況である。本発明においては、複数の話者の音声が個別に録音された音声データから、会話が行われている話者に対応した音声データの組み合せを判定することが可能となる音声処理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】
【０００５】
本発明が開示する音声処理装置は、第１音声が含まれる第１入力信号と、第２音声が含まれる第２入力信号を取得する取得部と、第１入力信号の第１信号強度と、第２入力信号声の第２信号強度を検出する検出部を備える。更に、当該音声処理装置は、第１信号強度の時系列と第２信号強度の時系列との相関係数を算出する算出部と、相関係数に基づいて、第１音声と第２音声が会話している状態か否かを判定する判定部を備える。
【０００６】
なお、本発明の目的及び利点は、例えば、請求項におけるエレメント及び組み合せにより実現され、かつ達成されるものである。また、上記の一般的な記述及び下記の詳細な記述の何れも、例示的かつ説明的なものであり、請求項のように、本発明を制限するものではないことを理解されたい。
【発明の効果】
【０００７】
本明細書に開示される音声処理装置では、複数の話者の音声が個別に録音された音声データから、会話が行われている話者に対応した音声データの組み合せを判定することが可能となる。
【図面の簡単な説明】
【０００８】
【図１】第１の実施形態による音声処理装置の機能ブロック図である。
【図２】音声処理装置による音声処理方法のフローチャートである。
【図３】一つの実施形態による検出部の機能ブロック図である。
【図４】実施例１に係る検出部３の検出処理のフローチャートである。

【図５】（ａ）は、検出部３による第１信号強度の第１検出結果を示す図である。（ｂ）は、検出部３による第２信号強度と第２信号強度の第２検出結果を示す図である。
【図６】（ａ）は、検出部３による第１信号強度の参考検出結果を示す図である。（ｂ）は、検出部３による第２信号強度と第２信号強度の参考検出結果を示す図である。
【図７】発話時間の比率の分布図である。
【図８】第１音声と第２音声が会話している状態における第１信号強度の時系列と第２信号強度の時系列との相関関係を示す図である。
【図９】第１信号強度の時系列と第２信号強度の時系列との相関係数の分布図である。
【図１０】実施例１に係る算出部４の算出処理と判定部５の判定処理のフローチャートである。
【図１１】実施例１と上述の比較例による複数の話者の音声が個別に録音された音声データから、会話をしている音声データの組み合わせの判定性能を示す図である。
【図１２】相関係数算出区間の概念図である。
【図１３】相関係数算出区間に含まれる話者交替回数と、会話している状態または会話していない状態における当該相関係数算出区間に基づいて算出した相関係数の関係を示す図である。
【図１４】信号強度の組み合わせに対する相関係数のテーブルを示す図である。
【図１５】実施例４に係る検出部３の検出処理と算出部４の算出処理のフローチャートである。
【図１６】実施例５に係る音声処理装置１の音声処理のフローチャートである。
【図１７】（ａ）は、検出部３による第２信号強度の第２検出結果を示す図である。（ｂ）は、検出部３による第３信号強度の第３検出結果を示す図である。（ｃ）は、検出部３による第４信号強度の第４検出結果を示す図である。
【図１８】一つの実施形態による音声処理装置として機能するコンピュータのハードウェア構成図である。
【発明を実施するための形態】
【０００９】
以下に、一つの実施形態による音声処理装置、音声処理方法及び音声処理プログラムの実施例を図面に基づいて詳細に説明する。なお、当該実施例は、開示の技術を限定するものではない。
【００１０】
（実施例１）
図１は、第１の実施形態による音声処理装置１の機能ブロック図である。音声処理装置１は、取得部２、検出部３、算出部４、判定部５を有する。図２は、音声処理装置１の音声処理のフローチャートである。実施例１においては、図２に示す音声処理装置１による音声処理のフローを、図１に示す音声処理装置１の機能ブロック図の各機能の説明に対応付けて説明する。
【００１１】
図１の取得部２は、例えば、ワイヤードロジックによるハードウェア回路である。また、取得部２は、音声処理装置１で実行されるコンピュータプログラムにより実現される機能モジュールであっても良い。取得部２は、入力音声の一例となる第１ユーザの第１音声と第２ユーザの第２音声を、例えば、外部装置を介して取得する。なお、当該処理は、図２に示すフローチャートのステップＳ２０１に対応する。取得部２は、第１音声が含まれる第１入力信号と第２音声が含まれる第２入力信号の取得方法として、例えば、特開２００８－１３４５５７号公報に記載の方法を用いることが出来る。ここで、第１ユーザと第２ユーザは、図示しないマイクロフォン（上述の外部装置に相当）をそれぞれ装着しているものとする。第１音声は、例えば、第１ユーザの会話の相手となるユーザ（第２ユーザまたは第２ユーザ以外のユーザ）に対して発話する音声を示す第１ユーザの音声である。また、第２音声は、例えば、第２ユーザの会話の相手となるユーザ（第１ユーザまたは第１ユーザ以外のユーザ）に対して発話する音声を示す第２ユーザの音声である。また、第

１音声と第２音声は、例えば、日本語であるが、英語等の他の言語であっても良い。換言すると、実施例１における音声処理においては、言語依存は存在しない。なお、実施例１においては、説明の便宜上、取得部２は第１ユーザの第１音声が含まれる第１入力信号と第２ユーザの第２音声が含まれる第２入力信号を取得するものとして説明するが、取得部２は第１ユーザまたは第２ユーザ以外の第ｘユーザ（ｘは、例えば、音声処理装置１を使用するユーザ数）の第ｘ音声が含まれる第x入力信号を取得しても良い。取得部２は取得した第１音声と第２音声を検出部３に出力する。

【００１２】

検出部３は、例えば、ワイヤードロジックによるハードウェア回路である。また、検出部３は、音声処理装置１で実行されるコンピュータプログラムにより実現される機能モジュールであっても良い。検出部３は、第１入力信号と第２入力信号を取得部２から受け取る。検出部３は、例えば、第１入力信号または第２入力信号に含まれる複数のフレームから、第１信号強度または第２信号強度を検出する。当該処理は、図２に示すフローチャートのステップＳ２０２に対応する。具体的には、検出部３は、第１入力信号に含まれる複数のフレーム毎の信号強度を示す第１信号強度を検出する。また、検出部３は、第２音声に含まれる複数のフレーム毎の信号強度を示す第２信号強度を検出する。なお、第１信号強度または第２信号強度は、例えば、第１音声または第２音声のパワーまたは信号対雑音比であれば良いが、当該パワーと信号対雑音比に限定されるものではない。

【００１３】

ここで、検出部３による第１入力信号または第２入力信号のパワーまたは信号対雑音比の検出処理の詳細について説明する。なお、説明の便宜上、第１音声のパワーを第１パワー、第１音声の信号対雑音比を第１信号対雑音比と称し、第２音声のパワーを第２パワー、第２音声の信号対雑音比を第２信号対雑音比と称するものとする。なお、第２パワーの検出方法は、第１パワーの検出方法と同様の手法を用いることができ、第２信号対雑音比の検出方法は、第１信号対雑音比の検出方法と同様の手法を用いることができる。この為、実施例１においては、検出部３による第１パワーと第１信号対雑音比の検出処理の詳細について説明する。

【００１４】

図３は、一つの実施形態による検出部３の機能ブロック図である。検出部３は、パワー算出部１０、雑音推定部１１、信号対雑音比算出部１２を有する。なお、検出部３は、パワー算出部１０、雑音推定部１１、信号対雑音比算出部１２を必ずしも有する必要はなく、各部が有する機能を、一つのまたは複数のワイヤードロジックによるハードウェア回路で実現させても良い。また、検出部３に含まれる各部が有する機能をワイヤードロジックによるハードウェア回路に代えて、音声処理装置１で実行されるコンピュータプログラムにより実現される機能モジュールで実現させても良い。

【００１５】

（検出部３によるパワー算出方法）

図３において、第１入力信号がパワー算出部１０に入力される。なお、パワー算出部１０は、図示しないバッファまたはキャッシュを有しても良い。パワー算出部１０は、第１入力信号に含まれる各フレームの音量を算出し、当該音量を雑音推定部１１と信号対雑音比算出部１２へ出力する。なお、第１入力信号に含まれる各フレーム長は、１６０サンプル（８ｋＨｚサンプリングで２０ｍｓｅｃに相当）である。第１音声信号の信号強度の一例となる第１パワーＰ１（ｔ）は、次式の通り、第１信号ｓ１（ｔ）のフレーム内の振幅二乗和に対して対数変換することで、算出することが出来る。

（数１）

$$Pl(t)=10\log\left(\sum_{i=t}^{t+L}{s_1}^2(i)\right)$$

上述の（数１）において、ｔはフレーム番号を示し、Ｌはフレーム長を示す。フレーム

長Ｌは、上述の通り、例えば１６０サンプル（８ｋＨｚサンプリングで２０ｍｓｅｃに相当）である。

【００１６】

（検出部３による信号対雑音比算出方法）

雑音推定部１１は、各フレームの第１パワーＰ１（ｔ）をパワー算出部１０から受け取る。雑音推定部１１は、各フレームにおける雑音を推定して、雑音推定結果を信号対雑音比算出部１２へ出力する。ここで、雑音推定部１１による各フレームの雑音推定は、例えば、以下の雑音推定方法を用いることが出来る。

【００１７】

（雑音推定方法）

雑音推定部１１は、第１入力信号のパワーＰ１（ｔ）と、１フレーム過去の第１雑音電力Ｎ１（ｔ－１）の差に応じて第１雑音電力を次式に基づいて更新する。雑音推定部１１は、第１音声信号のパワーＰ１（ｔ）と、１フレーム過去の第１雑音電力Ｎ１（ｔ－１）の差が所定閾値（例えば、５ｄＢ）を下回る場合は雑音パワーを音声信号が雑音であると推定し、第１雑音電力Ｎ１（ｔ）を更新する。一方で、雑音推定部１１は、第１入力信号のパワーＰ１（ｔ）と、１フレーム過去の第１雑音電力Ｎ１（ｔ－１）の差が所定閾値以上の場合は第１雑音電力を次式に基づいて更新しない。

（数２）

$$N1(t)=\begin{cases} N1(t-1)*COF+P1(t)*(1-COF) & ,\ (|P1(t)-N1(t-1)|<TH\_P\text{の場合}) \\ N1(n-1) & ,\ (\text{上記以外}) \end{cases}$$

上述の（数２）において、ＴＨ＿Ｐは雑音区間と判定するための判定閾値であり、例えば５ｄＢであれば良い。また、ＣＯＦは忘却係数であり、例えば０．０５であれば良い。雑音推定部１１は雑音推定結果となる各フレームの第１雑音電力（ｔ）を受け取る。

【００１８】

図３において、信号対雑音比算出部１２は、パワー算出部１０から各フレームの第１パワーＰ１（ｔ）を受け取り、雑音推定部１１から雑音推定結果となる各フレームの第１雑音電力Ｎ１（ｔ）を受け取る。なお、信号対雑音比算出部１２は、図示しないキャッシュまたはメモリを有しており、過去Ｌフレーム分の第１パワーＰ１（ｔ）、第１雑音電力Ｎ１（ｔ）を保持することができる。信号対雑音比算出部１２は、次式を用いて、第１信号対雑音比ＳＮＲ１（ｔ）を算出する。

（数３）

$$SNR1(t)=P1(t)-N1(t)$$

【００１９】

図４は、実施例１に係る検出部３の検出処理のフローチャートである。検出部３は、第１入力信号と第２入力信号を取得部２から受け取る（ステップＳ４０１）。次に、検出部３は、第１入力信号の第１パワーと第２入力信号の第２パワーを上述の（数１）に基づいて検出する（ステップＳ４０２）。次に、第１音声の第１雑音電力と第２音声の第２雑音電力を上述の（数２）に基づいて検出する（ステップＳ４０３）。そして、検出部３は、第１パワーと第２パワーを第１信号強度と第２信号強度として選択するか否かを決定する（ステップＳ４０４）。例えば、第１雑音電力と第２雑音電力がそれぞれ所定の閾値を下回る場合に、第１パワーと第２パワーを第１信号強度と第２信号強度として選択する。検出部３は、第１パワーと第２パワーを第１信号強度と第２信号強度として選択することを決定した場合（ステップＳ４０４－Ｙｅｓ）は、第１パワーと第２パワーを第１信号強度と第２信号強度として算出部４に出力する（ステップＳ４０７）ことで、図４のフローチャートに示す検出処理を終了する。一方、検出部３は、第１パワーと第２パワーを第１信号強度と第２信号強度として選択することを決定しない場合（ステップＳ４０４－Ｎｏ）は、検出部３は、第１音声の第１信号対雑音比と第２音声の第２信号対雑音比を上述の（

数３）に基づいて検出する（ステップＳ４０５）。次に、検出部３は、第１信号対雑音比と第２信号対雑音比を、第１信号強度または第２信号強度として算出部４に出力する（ステップＳ４０６）ことで、図４のフローチャートに示す検出処理を終了する。

【００２０】

図５（ａ）は、検出部３による第１信号強度の第１検出結果を示す図である。図５（ａ）の横軸は時間を示し、縦軸は第１信号強度の一例となる第１パワーを示している。図５（ｂ）は、検出部３による第２信号強度と第２信号強度の第２検出結果を示す図である。図５（ｂ）の横軸は時間を示し、縦軸は第２信号強度の一例となる第２パワーを示している。なお、図５（ａ）と図５（ｂ）の縦軸のパワーと横軸の時間は同一スケールである。検出部３は検出した第１信号強度と第２信号強度を算出部４に出力する。

【００２１】

図１において、算出部４は、例えば、ワイヤードロジックによるハードウェア回路である。また、算出部４は、音声処理装置１で実行されるコンピュータプログラムにより実現される機能モジュールであっても良い。算出部４は、第１信号強度と第２信号強度を検出部３から受け取る。算出部４は、第１信号強度と第２信号強度の時系列に対する相関係数を算出する。なお、当該処理は、図２に示すフローチャートのステップＳ２０３に対応する。具体的には、算出部４は、次式に基づいて、第１信号強度と第２信号強度の時系列に対する相関係数ｃｏｒｒ．を算出する。

（数４）

$$corr.=\frac{\sum_{i=Ts}^{Te}\left(P1'(i)-\overline{P1'(i)}\right)\left(P2(i)-\overline{P2(i)}\right)}{\sqrt{\sum_{i=Ts}^{Te}\left(P1'(i)-\overline{P1'(i)}\right)^2}\sqrt{\sum_{i=Ts}^{Te}\left(P2(i)-\overline{P2(i)}\right)^2}}$$

なお、上述の（数４）において、Ｔｓは相関係数を算出する始点時刻を示し、Ｔｅは相関係数を算出する終点時刻を示す。なお、Ｔｅ－Ｔｓは相関係数を算出する時間（時系列と称しても良い）となるが、任意の長さの時間を適用することが可能であり、例えば、６０秒であれば良い。算出部４は算出した第１信号強度と第２信号強度の時系列に対する相関係数を判定部５に出力する。

【００２２】

ここで、算出部４が、第１信号強度と第２信号強度の時系列に対する相関係数を算出する技術的意義について説明する。先ず、初めに本発明者らが検討した比較例について説明する。なお、当該比較例は公知技術ではなく、本発明者らにより新たに検証された事項であることを付言する。

【００２３】

（比較例）

複数の話者の音声が個別に録音された音声データに対し、会話が行われている話者に対応した音声データの組み合せを、２つの音声データの「発話時間の比率」に基づいて判定する方法を比較例として考える。図６（ａ）は、検出部３による第１信号強度の参考検出結果を示す図である。図６（ａ）の横軸は時間を示し、縦軸は第１信号強度の一例となる第１パワーを示している。図６（ｂ）は、検出部３による第２信号強度と第２信号強度の参考検出結果を示す図である。図６（ｂ）の横軸は時間を示し、縦軸は第２信号強度の一例となる第２パワーを示している。なお、図６（ａ）と図６（ｂ）の縦軸のパワーと横軸の時間は同一スケールである。ここで、任意の閾値（例えば、１０ｄＢ）以上の信号強度を満たす時間を発話時間と定義し、図６（ａ）と図６（ｂ）の場合における発話時間の比率（第１音声の発話時間／第２音声の発話時間）を算出すると、発話時間の比率は１．１となる。また、図５（ａ）と図５（ｂ）の場合における発話時間の比率を、図６（ａ）と図６（ｂ）の場合と同様に算出すると発話時間の比率は１．１となる。

【００２４】

ここで、自然な会話においては、話者同士が交互に発話を行うことでコミュニケーショ

ンを図る為、一方の話者の発話量が多い時間帯は、他方の話者の発話量が少ない（相手の発話に耳を傾ける）ことが想定される。ここで、図５（ａ）、図５（ｂ）と図６（ａ）、図６（ｂ）の組み合せを客観的に比較した場合、一方の話者の発話量が多い時間帯に、他方の話者の発話量が少ないのは、図５（ａ）、図５（ｂ）の組み合せであることが理解できる。図６（ａ）、図６（ｂ）の組み合せは、一方の話者の発話量が多い時間帯に、他方の話者の発話量も多くなっている為、一般的には第１ユーザと第２ユーザは、それぞれお互いに別のユーザと会話していることが想定される。

【００２５】

図７は、発話時間の比率の分布図である。図７においては、２０話者の音声データの組み合せとなる１９０組において、実際に会話をしている状態の１０組の音声データの組み合せと、会話を行っていない状態の１８０組の音声データの組み合せの発話時間の比率の分布図を示している。図７の発話時間の比率の分布図から理解できる通り、会話をしている音声データの組み合せの発話時間の比率と、会話を行っていない音声データの組み合せの発話時間の比率は同等程度であることが確認された。この為、複数の話者の音声が個別に録音された音声データから会話が行われている話者に対応した音声データの組み合せを、比較例に示した２つの音声データの「発話時間の比率」に基づいて判定することは難しいことが、本発明者らの検証により明らかになった。

【００２６】

一方、自然な会話においては、話者同士が交互に発話を行うことでコミュニケーションを図る為、図５（ａ）、図５（ｂ）に示される通り、一方の話者の発話量が多い時間帯は、他方の話者の発話量が少ないことが想定される。ここで、図５（ａ）、図５（ｂ）における、第１信号強度と第２信号強度の時系列に対する相関関係に着目すると、一方の変数（例えば、第１信号強度）が増大するほど、他方の変数（例えば、第２信号強度）が減少する「負の相関」の関係を有していることが理解できる。

【００２７】

この為、本実施例１の算出部４が、当該負の相関の強さを示す相関係数を算出することで、後述する判定部５が、当該相関係数に基づいて、第１音声と第２音声が会話している状態か否かを判定することができる。図８は、第１音声と第２音声が会話している状態における第１信号強度の時系列と第２信号強度の時系列との相関関係を示す図である。図８に示す通り、第１音声と第２音声が会話している状態においては、相関関係は負の相関（負の相関係数）を有することが理解できる。

【００２８】

図９は、第１信号強度の時系列と第２信号強度の時系列との相関係数の分布図である。図９においては、第１ユーザと第２ユーザの音声データを１１組（計３１分）用い、会話している区間と、会話していない区間（別の話者と会話している区間）を評価者が主観評価によって判定した上で、第１信号強度と第２信号強度の時系列に対する相関係数をそれぞれ算出した。図９から理解できる通り、会話している区間と会話していない区間で相関係数の範囲が明確に異なっていることが確認された。図９の実験結果より、任意の閾値である第１閾値を例えば、－０．４（一般的に相関の強さが中程度の値）と規定し、当該第１閾値と第１信号強度と第２信号強度の時系列に対する相関係数を用いることで、第１音声と第２音声が会話している状態か否（会話していない状態）かを判定することができる。

【００２９】

図１において、判定部５は、例えば、ワイヤードロジックによるハードウェア回路である。また、判定部５は、音声処理装置１で実行されるコンピュータプログラムにより実現される機能モジュールであっても良い。判定部５は、算出部４が算出した第１信号強度の時系列と第２信号強度の時系列との相関係数を算出部４から受け取る。判定部５は、相関係数に基づいて第１音声と第２音声が会話している状態か否かを判定する。なお、当該処理は、図２に示すフローチャートのステップＳ２０４に対応する。判定部５は、相関係数が負であり、相関係数が上述の第１閾値未満（例えば、－０．４＞相関係数＞＝－１．０

）の場合に第１音声と第２音声が会話している状態であると判定する。換言すると、判定部５は、相関係数が第１閾値以上（例えば、－０．４＜＝相関係数＜＝＋１．０）の場合に、第１音声と第２音声が会話していない状態であると判定しても良い。判定部５は、判定結果を任意の外部装置（例えば、図示しないコンピュータ）に出力する。

【００３０】

判定部５は、必要に応じて、第１音声と第２音声が会話している状態と判定した相関係数の算出に用いられた、上述の（数４）におけるＴｅ－Ｔｓの区間長を算出部４から受け取り、当該区間長を、第１ユーザと第２ユーザが会話した区間（時間）として外部装置に出力することもできる。または、判定部５は、必要に応じて、第１音声と第２音声が会話している状態と判定した相関係数の算出に用いられた、上述の（数４）における始点時刻のＴｓと、終点時刻のＴｅを算出部４から受け取り、当該Ｔｓを第１ユーザと第２ユーザの会話の始点時刻、Ｔｅを第１ユーザと第２ユーザの会話の終点時刻として外部装置に出力することもできる。

【００３１】

図１０は、実施例１に係る算出部４の算出処理と判定部５の判定処理のフローチャートである。算出部４は、第１信号強度と第２信号強度を検出部３から受け取る（ステップＳ１００１）。次に、算出部４は、第１信号強度と第２信号強度の時系列に対する相関係数を（数４）に基づいて算出する（ステップＳ１００２）。算出部４は、算出した相関係数を判定部５に出力する（ステップＳ１００３）ことで、図１０のフローチャートに示す算出部４の算出処理を終了する。

【００３２】

判定部５は、算出部４が算出した相関係数を算出部４から受け取る（ステップＳ１００４）。判定部５は、相関係数が上述の第１閾値未満か否かを判定する（ステップＳ１００５）。相関係数が上述の第１閾値未満の場合（ステップＳ１００５－Ｙｅｓ）、判定部５は、第１音声と第２音声が会話している状態として判定し（ステップＳ１００６）、その判定結果を任意の外部装置に出力する（ステップＳ１００８）ことで、図１０のフローチャートに示す判定部５の判定処理を終了する。一方、相関係数が上述の第１閾値以上の場合（ステップＳ１００５－Ｎｏ）、判定部５は、第１音声と第２音声が会話していない状態として判定し（ステップＳ１００７）、その判定結果を任意の外部装置に出力する（ステップＳ１００８）ことで、図１０のフローチャートに示す判定部５の判定処理を終了する。

【００３３】

図１１は、実施例１と上述の比較例による複数の話者の音声が個別に録音された音声データから、会話をしている音声データの組み合せの判定性能を示す図である。図１１は、２０話者の音声データの組み合せ計１９０組（会話している音声データの組み合せ１０組、会話していない音声データの組み合せ１８０組の合計）において、会話している音声データの検出率（会話している音声を検出した割合）と、当該検出した会話している音声データの正解率（検出した音声が会話している状態である割合）を示している。なお、比較例においては、検出率が１００％となるように発話時間比が１．２以下の音声データの組み合せを会話しているデータとして判定した。図１１から理解できる通り、実施例１によれば、比較例に対して正解率が大幅に向上していることが確認された。実施例１における音声処理装置１に依れば、会話において一方が発話している場合、他方は発話しない特徴を利用し、異なるユーザ同士の音声信号強度の時系列に対する相関係数が負であり、かつ相関係数が上述の第１閾値未満の場合にユーザ同士が会話していると判定することが可能となる。換言すれば、実施例１における音声処理装置１は、複数の話者の音声が個別に録音された音声データから、会話が行われている話者に対応した音声データの組み合せを、高い精度で判定することが可能となる。

【００３４】

（実施例２）

図１の検出部３は、実施例１の検出処理に加えて、第１信号強度と第２信号強度の時系

列に対する大小関係を検出し、当該大小関係が反転する反転回数が第４閾値以上（例えば、第４閾値＝６回）となる相関係数算出区間を検出しても良い。また、図１の算出部４は、実施例１の算出処理に加えて、実施例２に係る検出部３が検出する相関係数算出区間における相関係数を算出しても良い。以下、実施例２に係る音声処理装置１の音声処理の詳細について説明する。

【００３５】

検出部３は、第１信号強度と第２信号強度の時系列に対する大小関係を検出する。具体的には、検出部３は、第１信号強度と第２信号強度（例えば、第１パワーと第２パワー）の差が所定の第５閾値（例えば、パワー差＝＋２０ｄＢ）以上となる第１状態と、第１信号強度と第２信号強度（例えば、第１パワーと第２パワー）の差が所定の第６閾値（例えば、パワー差＝－２０ｄＢ）以下となる第２状態を検出する。検出部３は、第１状態から第２状態へ遷移した場合、または、第２状態から第１状態に遷移した場合を話者交代点（第１信号強度と第２信号強度の時系列に対する大小関係が反転する点と称しても良い）として検出する。検出部３が、話者交代点が所定の第４閾値（例えば、第４閾値＝６回）となる相関係数算出区間を検出する。図１２は、相関係数算出区間の概念図である。図１２の横軸は第１音声または第２音声の時系列であり、縦軸は、第１信号強度と第２信号強度の差の一例となるパワー差である。なお、図１２のパワー差は、図５（ａ）と図５（ｂ）のパワー差を示している。検出部３は、話者交代点が第４閾値以上を満たす区間を相関係数算出区間として検出する。検出部３は、検出した相関係数算出区間を算出部４に出力する。

【００３６】

算出部４は、検出部３が検出した相関係数算出区間を検出部３から受け取る。算出部４は、相関係数算出区間における第１信号強度と第２信号強度の時系列に対する相関係数を実施例１と同様の手法を用いて算出する。算出部４は、上述の（数４）のＴｓを相関係数算出区間の始点時刻とし、Ｔｅを相関係数算出区間の終点時刻とし、Ｔｅ－Ｔｓで表現される相関係数算出区間における第１信号強度と第２信号強度の時系列に対する相関係数を算出すれば良い。なお、算出部４は、検出部３から複数の相関係数算出区間を受け取っている場合は、複数の相関係数算出区間毎に相関係数を算出すれば良い。算出部４は、算出した相関係数を判定部５に出力し、判定部５は実施例１と同様の方法で判定処理を行えば良い。なお、判定部５は、複数の相関係数を算出部４から受け取っている場合は、実施例１と同様の方法で相関係数毎に判定処理を行えば良い。

【００３７】

ここで、実施例２における技術的意義について説明する。図１３は、相関係数算出区間に含まれる話者交替回数と、会話している状態または会話していない状態における当該相関係数算出区間に基づいて算出した相関係数の関係を示す図である。図１３から理解できる通り、話者交代回数が多い区間（例えば、上述の第４閾値となる６回）の場合、会話している音声データの組み合せの相関係数と、会話していない音声データの組み合せの差（有意差）が現れる。この為、実施例２における音声処理装置１においては、話者交代回数が所定の第４閾値以上となる相関係数算出区間で相関係数を算出することで、高い精度で複数の話者の音声が個別に録音された音声データから、会話が行われている話者に対応した音声データの組み合せを、高い精度で判定することが可能となる。なお、実施例２における音声処理装置１は、実施例１に記載した音声処理を任意に組み合せることができる。

【００３８】

（実施例３）

図１の算出部４は、実施例１または実施例２の算出処理に加えて、第１信号強度の第１位相または、第２信号強度の第２位相を所定の範囲で変化させて複数の相関係数を算出しても良い。また、図１の判定部５は、算出部４が第１位相または第２位相を所定の範囲で変化させて算出した複数の相関係数の中で、最小値を満たす相関係数に基づいて、第１音声と第２音声が会話している状態か否かを判定しても良い。以下、実施例３に係る音声処理装置１の音声処理の詳細について説明する。

【００３９】
算出部４は、第１信号強度の第１位相または、第２信号強度の第２位相を所定の範囲で変化させて複数の相関係数を算出する。実施例３においては、第１位相を所定の範囲で変化させた場合について説明するが、第２位相を所定の範囲で変化させる場合も第１位相の場合と同様に算出することが出来る為、第２位相に関する詳細な説明は省略する。算出部４は、次式に基づいて、第１信号強度（例えば、第１パワーＰ１）の位相を変化させた場合の第１信号強度Ｐ１（ｔ）と第２信号強度Ｐ２（ｔ）の複数の相関係数から、最小値の相関係数を満たす位相ｄｍｉｎを次式に基づいて算出する。
（数５）

$$dmin = \underset{d}{\operatorname{argmin}} \left\{ \frac{\sum_{i=Ts}^{Te}\left(P1(i)+d-\overline{P1(i)}\right)\left(P2(i)-\overline{P2(i)}\right)}{\sqrt{\sum_{i=Ts}^{Te}\left(P1(i)+d-\overline{P1(i)}\right)^2}\sqrt{\sum_{i=Ts}^{Te}\left(P2(i)-\overline{P2(i)}\right)^2}} \right\} \mid \mathrm{d}=\text{-Dmax}, \cdots, Dmax$$

$$\overline{P1(i)} = \frac{1}{Te-Ts+1}\sum_{i=Ts}^{Te} P1(i)$$

上述の（数５）においてｄは位相の変更量（サンプル）を示し、Ｄｍａｘは位相変更の最大値を示す。Ｄｍａｘは、例えば８００００サンプル（１０秒に相当）とすれば良い。算出部４は、第１強度の位相を所定の範囲となるｄｍｉｎ変化（シフトと称しても良い）させた信号Ｐ１’ｔを次式に基づいて算出する。
（数６）
Ｐ１’（ｔ）＝Ｐ１（ｔ＋ｄｍｉｎ）

判定部５は、複数の相関係数のうち最小値の満たす相関係数に基づいて、第１音声と第２音声が会話している状態か否かを実施例１と同様の判定方法を用いて判定する。
【００４０】
実施例３における音声処理装置１によれば、例えば、取得部２が取得する第１入力信号と第２入力信号のそれぞれにおいて、録音のタイミングが同期していない場合（例えば、第１ユーザと第２ユーザがそれぞれ装着しているマイクロフォンの音声処理の内部処理の差異や、設定時刻のずれ等に起因して録音のタイミングが同期していない場合）においても、相関係数を算出する時点において、位相を調整することによって、会話している音声データの組み合せの判定精度を向上させることが可能となる。なお、実施例３における音声処理装置１は、実施例１または実施例２に記載した音声処理を任意に組み合せることができる。
【００４１】
（実施例４）
図１の検出部３は、実施例１ないし実施例３の検出処理に加え、第１入力信号に含まれる第１発話区間を第１信号強度に基づいて検出し、第２入力信号に含まれる第２発話区間を第２信号強度に基づいて検出し、第１発話区間と第２発話区間が重複する重複発話区間を検出しても良い。また、図１の算出部４は、実施例１ないし実施例３の算出処理に加え、当該重複発話区間が第２閾値未満の第１信号強度と第２信号強度を、相関係数の算出に用いない算出処理を行っても良い。また、検出部３は、第１入力信号に含まれる第１無音区間を第１信号強度に基づいて検出し、第２入力信号に含まれる第２無音区間を第２信号強度に基づいて検出し、第１無音区間と第２無音区間が重複する重複無音区間を検出しても良い。算出部４は、当該重複無音区間が第３閾値未満の第１信号強度と第２信号強度を、相関係数の算出に用いない算出処理を行っても良い。以下、実施例４に係る音声処理装置１の音声処理の詳細について説明する。
【００４２】

検出部３は、例えば、第１信号のフレーム単位で、第１信号強度が所定の閾値（例えば、１０ｄＢ（但し、第１信号強度として第１パワーを用いる場合））以上か否かを判定する。検出部３は、第１信号強度が当該閾値以上を満たすフレームを第１発話区間として判定する。また、検出部３は、第１信号強度が当該閾値未満を満たすフレームを第１無音区間として判定する。検出部３は、次式に基づいて、フレーム単位で当該フレームが第１発話区間か第１無音区間（非発話区間と称しても良い）を判定し、判定結果ｖ１（ｔ）を検出する。

（数７）

$$v_1(t)=1 \quad (\text{第１発話区間})$$
$$v_1(t)=0 \quad (\text{第１無音区間})$$

上述の（数７）において、ｔはフレーム番号を示す。なお、１フレームの長さは、例えば、２０ｍｓである。また、上述の（数７）においては、tフレーム目の第１音声が発話区間と判定された場合はｖ１（ｔ）＝１が代入され、tフレーム目の第１音声が無音区間と判定された場合はｖ１（ｔ）＝０と代入されることを意味する。検出部３は、ｖ１（ｔ）＝１を連続して満たすフレーム区間を第１発話区間として出力する。なお、検出部３は、第２音声に関する判定結果ｖ２（ｔ）をｖ１（ｔ）と同様の手法を用いて算出し、第２発話区間または第２無音区間を検出する。

【００４３】

次に、検出部３は、第１発話区間と第２発話区間が重複する重複発話区間を検出する。当該重複発話区間は、例えば、ある任意の同一の時刻において、第１ユーザと第２ユーザが互いに発話している区間と定義することが出来る。なお、検出部３は、具体的には、次式に基づいて重複発話区間ＴＯ（ｔ）を検出することが出来る。

（数８）

$$\textit{if} \quad \{v_1(t)=0\}\vee\{v_2(t)=0\} \qquad TO(t)=0$$
$$\textit{else} \qquad TO(t)=TO(t-1)+1$$

上述の（数８）は、第１ユーザの第１音声と第２ユーザの第２音声の何れかが無音区間と判定されるフレームに対しては重複時間を０（重複区間の発現無し）と規定し、第１ユーザの第１音声と第２ユーザの第２音声の双方が発話区間と判定されるフレームに対しては直前のフレームまでの重複時間に１フレーム加算することで、重複が連続するフレーム数（重複区間）を算出することを意味する。検出部４は、規定した重複発話区間を算出部５に出力する。なお、重複発話区間には、区間の長さの情報が含まれているものとする。重複発話区間の長さＬＯは、例えば、次式に基づいて算出することができる。

（数９）

ＬＯ＝ＴＯ＿ｅ（ｉ）－ＴＯ＿ｓ（ｉ）

なお、上述の（数９）において、ＴＯ＿ｓ（ｉ）は、重複発話区間の始点（開始フレーム）であり、ＴＯ＿ｅ（ｉ）は、重複発話区間の終点（終了フレーム）である。

【００４４】

次に、検出部３は、第１無音区間と第２無音区間が重複する重複無音区間を検出する。当該重複無音区間は、例えば、ある任意の同一の時刻において、第１ユーザと第２ユーザが互いに発話していない区間と定義することが出来る。なお、検出部３は、具体的には、次式に基づいて重複無音区間ＴＥ（ｔ）を検出することが出来る。

（数１０）

$$if \quad \{v_1(t)=1\}\vee\{v_2(t)=1\} \quad or\{v_1(t)=1\}\vee\{v_2(t)=0\} \quad or\{v_1(t)=0\}\vee\{v_2(t)=1\} \quad TE(t)=0$$
$$else \qquad TE(t)=TE(t-1)+1$$

上述の（数１０）は、第１ユーザの第１音声と第２ユーザの第２音声の何れかまたは双方が発話区間と判定されるフレームに対しては重複時間を０（重複区間の発現無し）と規定し、第１ユーザの第１音声と第２ユーザの第２音声の双方が無音区間と判定されるフレームに対しては直前のフレームまでの重複時間に１フレーム加算することで、重複が連続するフレーム数（重複区間）を算出することを意味する。検出部４は、規定した重複無音区間を算出部４に出力する。なお、重複無音区間には、区間の長さの情報が含まれているものとする。重複無音区間の長さＬＳは、例えば、次式に基づいて算出することができる。

（数１１）

ＬＳ＝ＴＳ＿ｅ（ｉ）－ＴＳ＿ｓ（ｉ）

なお、上述の（数１１）において、ＴＳ＿ｓ（ｉ）は、重複無音区間の始点（開始フレーム）であり、ＴＳ＿ｅ（ｉ）は、重複無音区間の終点（終了フレーム）である。

【００４５】

（重複発話区間に対する処理）

算出部４は、重複発話区間の長さが所定の第２閾値未満（例えば、第２閾値＝１秒（５０フレームに相当））となる第１信号強度と第２信号強度を相関係数の算出に用いない算出処理を行うことができる。換言すると、算出部４は、重複発話区間が第２閾値未満となる第１信号強度と第２信号強度以外の相関係数を用いて相関係数の算出処理を実施例１と同様の方法で行う。ここで、第２閾値未満の重複発話区間の長さは、例えば、「はい」、「いいえ」、「うんうん」、「本当？」、「そうです」等の一般的に相槌の長さに相当する。相槌は、実際に会話している状態において発生するものである。一方が発話している時間帯に他方も発話する（相槌を打つ）場合も存在する為、重複発話区間の長さが第２閾値未満となる第１信号強度と第２信号強度を相関係数の算出に用いないことで、相槌の影響を排除することが可能となる。なお、重複発話区間の長さが第２閾値未満となる第１信号強度と第２信号強度を相関係数の算出に用いた場合、互いに会話している状態であるにも係らず、第１信号強度と第２信号強度が、互いにある程度の強さの信号強度を有している為、相関係数は会話していない状態を示す正の値に近づくことになる。

【００４６】

（重複無音区間に対する処理）

算出部４は、重複無音区間の長さが所定の第３閾値未満（例えば、第３閾値＝１０秒（５００フレームに相当））となる第１信号強度と第２信号強度を相関係数の算出に用いない算出処理を行うことができる。換言すると、算出部４は、重複発話区間が第３閾値未満となる第１信号強度と第２信号強度以外の相関係数を用いて相関係数の算出処理を実施例１と同様の方法で行う。ここで、第３閾値未満の重複無音区間の長さは、会話している状態で発生することが想定される無音区間である。実際に会話している場合においては、例えば、双方が発話する内容を考えることより、無音区間が重複する場合も想定される。この為、算出部４が、重複発話区間の長さが第３閾値未満となる第１信号強度と第２信号強度を相関係数の算出に用いないことで、上述の影響を排除することが可能となる。

【００４７】

図１５は、実施例４に係る検出部３の検出処理と算出部４の算出処理のフローチャートである。検出部３は、第１音声の第１発話区間と第１無音区間を、上述の（数７）に基づいて検出する（ステップＳ１５０１）。次に、検出部３は、第２音声の第２発話区間と第２無音区間を、上述の（数７）に基づいて検出する（ステップＳ１５０２）。次に、検出部３は、第１発話区間と第２発話区間から重複発話区間を上述の（数８）に基づいて検出する（ステップＳ１５０３）。次に、検出部３は、第１無音区間と第２無音区間から重複無音区間を上述の（数１０）に基づいて検出する（ステップＳ１５０４）。検出部３は、

検出した重複発話区間と重複無音区間を算出部４に出力することで、図１５のフローチャートに示す検出処理を終了する。

【００４８】

算出部４は、重複発話区間と重複無音区間を検出部３から受け取り、重複発話区間が上述の第２閾値未満、かつ重複無音区間が上述の第３閾値未満か否かを判定する（ステップＳ１５０５）。重複発話区間が上述の第２閾値未満、かつ重複無音区間が上述の第３閾値未満の場合（ステップＳ１５０５－Ｙｅｓ）、算出部４は、重複発話区間と重複無音区間を用いないで相関係数を上述の（数４）に基づいて算出し（ステップＳ１５０９）、算出した相関係数を判定部５に出力する（ステップＳ１５１２）ことで、図１５のフローチャートに示す算出処理を終了する。

【００４９】

重複発話区間が上述の第２閾値未満、かつ重複無音区間が上述の第３閾値未満ではない場合（ステップＳ１５０５－Ｎｏ）、算出部４は、重複発話区間が第２閾値未満、かつ重複無音区間が第３閾値以上か否かを判定する（ステップＳ１５０６）。重複発話区間が第２閾値未満、かつ重複無音区間が第３閾値以上の場合（ステップＳ１５０６－Ｙｅｓ）、算出部４は、重複発話区間を用いないで相関係数を上述の（数４）に基づいて算出し（ステップＳ１５１０）、算出した相関係数を判定部５に出力する（ステップＳ１５１２）ことで、図１５のフローチャートに示す算出処理を終了する。

【００５０】

重複発話区間が上述の第２閾値未満、かつ重複無音区間が上述の第３閾値以上ではない場合（ステップＳ１５０６－Ｎｏ）、算出部４は、重複発話区間が第２閾値以上、かつ重複無音区間が第３閾値未満か否かを判定する（ステップＳ１５０７）。重複発話区間が第２閾値以上、かつ重複無音区間が第３閾値未満の場合（ステップＳ１５０７－Ｙｅｓ）、算出部４は、重複無音区間を用いないで相関係数を上述の（数４）に基づいて算出し（ステップＳ１５１１）、算出した相関係数を判定部５に出力する（ステップＳ１５１２）ことで、図１５のフローチャートに示す算出処理を終了する。

【００５１】

重複発話区間が第２閾値以上、かつ重複無音区間が第３閾値未満ではない場合（ステップＳ１５０７－Ｎｏ）、算出部４は、実施例１と同様の方法で、相関係数を上述の（数４）に基づいて算出し（ステップＳ１５０８）、算出した相関係数を判定部５に出力する（ステップＳ１５１２）ことで、図１５のフローチャートに示す算出処理を終了する。

【００５２】

実施例４においては、説明の便宜上、重複発話区間と重複無音区間に対する双方の処理について説明したが、重複発話区間または重複無音区間の何れか一方のみを処理対象としても良いし、双方を処理対象としても良い。実施例４に係る音声処理装置においては、複数の話者の音声が個別に録音された音声データから、会話が行われている話者に対応した音声データの組み合せを、重複発話区間または重複無音区間を考慮することで、高い精度で判定することが可能となる。なお、実施例４における音声処理装置１は、実施例１ないし実施例３に記載した音声処理を任意に組み合せることができる。

【００５３】

（実施例５）

図１の取得部２は、第１ユーザの第１音声が含まれる第１入力信号または第２ユーザの第２音声が含まれる第１入力信号のみならず、複数のユーザ（第ｘユーザと称しても良い）の各音声が含まれる各入力信号（第ｘ入力音声と称しても良い）を取得することが出来る。実施例５においては、複数のユーザの各音声の組み合せにおいて、実際に会話をしている２つの音声データの組み合せを判定する方法について説明する。なお、実施例５は、説明の便宜上、第１ユーザの第１音声ないし第４ユーザの第４音声を取得部２が外部装置を介して取得するものとして説明する。ここで、第１ユーザないし第４ユーザは、図示しないマイクロフォン（上述の外部装置に相当）をそれぞれ装着しているものとする。取得部２は取得した第１音声ないし第４音声を検出部３に出力する。

【００５４】
図１の検出部３は、第１入力音声ないし第４入力音声を取得部３から受け取る。検出部３は、第１信号ないし第４音声の各信号強度（第１信号強度、第２信号強度、第３信号強度、第４信号強度）を、例えば、実施例１に記載した検出方法を用いて検出する。検出部３は、検出した第１信号強度ないし第４信号強度を算出部４に出力する。
【００５５】
図１の算出部４は、各信号強度、換言すると第１信号強度ないし第４信号強度を検出部３から受け取る。算出部４は、各信号強度から２つの信号強度を組み合せた場合の時系列に対する各相関係数を算出する。具体的には、算出部４は、第１信号強度ないし第４信号強度から２つ信号強度を組み合せる。図１４は、信号強度の組み合せに対する相関係数のテーブルを示す図である。算出部４は、例えば、算出部４が有する図示しないバッファまたはメモリに図１４に示す相関係数のテーブルを格納し、音声データの組み合せＩＤに対応付けて第１信号強度ないし第４信号強度から２つ信号強度を組み合せる。算出部４は、図１４の組み合せＩＤに対応付けられている、２つの信号強度を組み合せた場合の、当該２つの信号強度の時系列に対する各相関係数を、例えば、実施例１で記載した算出方法を用いて算出する。算出部４は、例えば、図１４の相関係数のテーブルに、各相関係数を格納しても良い。算出部４は、算出した各相関係数を判定部５に出力する。
【００５６】
図１の判定部５は、第１信号強度ないし第４信号強度から２つの信号強度を組み合せた場合の時系列に対する各相関係数を算出部４から受け取る。判定部５は、各相関係数のうち、相関係数が最小値を満たす２つの信号強度の組み合せに基づいて、複数のユーザの中から会話しているユーザの組み合せを判定する。例えば、図１４の相関係数の組み合せテーブルを参照すると、組み合せＩＤ４における第２信号強度と第３信号強度が最小値を満たす相関係数である為、判定部５は、第２ユーザと第３ユーザが会話している状態のユーザの組み合せとして判定する。また、判定部５は、第２信号強度と第３信号強度が最小値となる場合において、第２信号強度と第３信号強度の時系列に対する相関係数が上述の第１閾値未満の場合に、組み合せた２つの音声データが会話している状態として判定しても良い。
【００５７】
図１６は、実施例５に係る音声処理装置１の音声処理のフローチャートである。取得部２は、第１音声ないし第４音声（複数の音声と称しても良い）を取得する（ステップＳ１６０１）。次に、検出部３は、第１信号強度ないし第４信号強度を検出する（ステップＳ１６０２）。次に、検出部３は、第１信号強度ないし第４信号強度から２つの信号強度を組み合せる（ステップＳ１６０３）。次に、算出部４は、組み合せた２つの信号強度の時系列に対する相関係数を（数４に）基づいて算出する（ステップＳ１６０４）。次に、算出部４は、組み合せた全ての２つの信号強度の時系列に対する相関係数を算出したか否かを判定する（ステップＳ１６０５）、組み合せた全ての２つの信号強度の時系列に対する相関係数を算出していない場合（ステップＳ１６０５－Ｎｏ）、算出部４は、ステップＳ１６０４とＳ１６０５の処理を繰り返し実施する。組み合せた全ての２つの信号強度の時系列に対する相関係数を算出している場合（ステップＳ１６０５－Ｙｅｓ）、算出部４は、相関係数が最小値を満たす２つの信号強度を判定する（ステップＳ１６０６）。
【００５８】
次に、判定部５は、最小値を満たす相関係数が上述の第１閾値未満か否かを判定する（ステップＳ１６０７）。相関係数が上述の第１閾値未満の場合（ステップＳ１６０７－Ｙｅｓ）、判定部５は、相関係数が最小値を満たす２つの信号強度が会話している状態として判定し（ステップＳ１６０８）、その判定結果を任意の外部装置に出力する（ステップＳ１６１０）ことで、音声処理装置１は、図１６のフローチャートに示す音声処理を終了する。相関係数が上述の第１閾値以上の場合（ステップＳ１６０７－Ｎｏ）、判定部５は、組み合せた２つの信号強度の何れも会話していない状態として判定し、（ステップＳ１６０９）、その判定結果を任意の外部装置に出力する（ステップＳ１６１０）ことで、音

声処理装置１は、図１６のフローチャートに示す音声処理を終了する。

【００５９】

実施例５における音声処理装置１によれば、複数の話者の音声が個別に録音された音声データから、会話が行われている話者に対応した音声データの組み合せを、相関係数が最小値を満たす２つの信号強度の組み合せに基づいて判定することで、高い精度で判定することが可能となる。なお、実施例５における音声処理装置１は、実施例１ないし実施例４に記載した音声処理を任意に組み合せることができる。

【００６０】

（実施例６）

実施例６における音声処理装置１においては、３人以上で会話しているグループを特定することが可能である。例えば、職場等においては、会話している話者は２名とは限らず、３名以上のグループで会話をしている場合も存在する。以下、３人以上で会話しているグループを特定する音声処理装置１の音声処理の詳細について説明する。実施例６における取得部２、検出部３の処理、ならびに算出部４の一部の処理は、実施例５と同様の処理である為、重複する処理に関する詳細な説明は省略する。

【００６１】

図１の判定部５は、第１信号強度ないし第４信号強度から２つの信号強度を組み合せた場合の時系列に対する各相関係数を算出部４から受け取る。判定部５は、実施例５と同様に、各相関係数のうち、相関係数が最小値を満たす２つの信号強度の組み合せに基づいて、複数のユーザの中から会話している信号強度の組み合せを判定する。例えば、図１４の相関係数の組み合せテーブルを参照すると、組み合せＩＤ４における第２信号強度と第３信号強度が最小値を満たす相関係数の組み合せになる。次に、算出部４は、最小値を満たす相関係数の算出に用いた２つの信号強度を加算した加算信号強度を算出する。具体的には、例えば、算出部４は第２信号強度Ｐ２（ｔ）と第３信号強度Ｐ３（ｔ）を加算した加算信号強度Ｐａｄｄ（ｔ）を次式（（数１２）または（数１３）の何れか一方、または双方の平均値を用いても良い）に基づいて算出する。

（数１２）

$$P\_ADD_n(t) = \sum_i P_i(t) \,\middle|\, i \in G_n$$

（数１３）

$$P\_ADD_n(t) = \max_i P_i(t) \,\middle|\, i \in G_n$$

上述の（数１２）と（数１３）において、ｎは第ｘユーザに対応する。また、上述の（数１２）は、信号強度の合計値を示し、（数１３）は信号強度の最大値を示す。また、Ｇｎはグループを示す。例えば、第１信号と第２信号がグループに含まれる場合は、Ｇｎ＝｛１、２｝と表現され、第２信号と第３信号がグループに含まれる場合は、Ｇｎ＝｛２、３｝と表現される。

【００６２】

次に、算出部４は、加算信号強度と、最小値を満たす相関係数の算出に用いた２つの信号強度以外（例えば、第２信号強度または第３信号強度以外）１つの信号強度（第１信号強度または第４信号強度）との時系列に対する参照相関係数を算出する。実施例６においては、説明の便宜上、加算信号強度と第４ユーザの第４信号強度の時系列に対する参照相関係数ｃｏｒｒ．$_n$を次式に基づいて算出する。

（数１４）

$$corr_n = \frac{\sum_{i=Ts}^{Te}\left(P1(i)-\overline{P1(i)}\right)\left(P\_ADDn(i)-\overline{P\_ADD(i)}\right)}{\sqrt{\sum_{i=Ts}^{Te}\left(P1(i)-\overline{P1(i)}\right)^2}\sqrt{\sum_{i=Ts}^{Te}\left(P\_ADD(i)-\overline{P\_ADD(i)}\right)^2}}$$

上述の（数１４）において、実施例６においては、Ｐ１（ｉ）は、第４信号強度であり、Ｐａｄｄ（ｔ）は、加算信号強度である。なお、算出部４は、最小値を満たす相関係数の算出に用いた２つの信号強度の重複無音区間が第５閾値以上（例えば、重複無音区間＝１０秒）の場合に、参照相関係数を算出しても良い。算出部４は、２つの信号強度の組み合せの中で最小値を満たす相関係数と、参照相関係数を判定部５に出力し、判定部５はそれらを受け取る。

【００６３】

判定部５は、参照相関係数と２つの信号強度の組み合せの中で最小値を満たす相関係数を算出部４から受け取る。判定部５は、参照相関係数が最小値を満たす相関係数未満の場合、または、参照相関係数が上述の第１閾値未満の場合に、参照相関係数の算出に用いた３つの信号強度に基づいて、会話している３人以上のユーザの組み合せ（会話しているグループと称しても良い）を判定する。例えば、判定部５は、参照相関係数が最小値を満たす相関係数未満の場合、第２ユーザと第３ユーザと第４ユーザが３名のグループで会話しているものと判定する。また、判定部５は、参照相関係数が、最小値を満たす相関係数未満の場合、かつ、第１閾値未満の場合、第２ユーザと第３ユーザと第４ユーザが３名のグループで会話しているものと判定しても良い。更に、判定部５は、参照相関係数が、最小値を満たす相関係数以上の場合であっても、第１閾値未満の場合であれば、第２ユーザと第３ユーザと第４ユーザが３名のグループで会話しているものと判定しても良い。なお、例えば、上述の（数１２）または（数１３）において、グループＧｎ＝｛２、３｝であり（第２信号と第３信号がグループに含まれる場合）、第４音声を加える場合は、新たなグループとして、Ｇｎ＋１＝ＧｎＵ｛３｝＝｛２、３、４｝が定義されれば良い。

【００６４】

また、判定部５が、第２ユーザと第３ユーザと第４ユーザが３名のグループで会話しているか否かを判定した後、以下の処理を実施することで、第１ユーザを含めた４名のグループで会話しているか否かを判定することができる。先ず、算出部４は、第２信号強度、第３信号強度、第４信号強度に基づいて算出した参照相関係数を、最小値を満たす相関係数に置換する。次に、算出部４は、第１信号強度、第２信号強度、ならびに第３信号強度の上述の（数１２）または、（数１３）に基づいて、加算信号強度を算出する。次に、次に、算出部４は、第１信号強度と、加算信号強度の時系列に対する参照相関係数を算出する。判定部５の処理は、実施例６に開示している同様の判定方法を用いれば良い。判定部５は、全ての信号強度に対して上述の処理を行い、会話をしている複数のユーザ（グループ）を判定することもできる。

【００６５】

ここで、実施例６の技術的意義について説明する。図１７（ａ）は、検出部３による第２信号強度の第２検出結果を示す図である。図１７（ａ）の横軸は時間を示し、縦軸は第２信号強度の一例となる第２パワーを示している。図１７（ｂ）は、検出部３による第３信号強度の第３検出結果を示す図である。図１７（ｂ）の横軸は時間を示し、縦軸は第３信号強度の一例となる第３パワーを示している。図１７（ｃ）は、検出部３による第４信号強度の第４検出結果を示す図である。図１７（ｃ）の横軸は時間を示し、縦軸は第４信号強度の一例となる第４パワーを示している。なお、図１７（ａ）ないし図１７（ｃ）の縦軸のパワーと横軸の時間は同一スケールである。

【００６６】

２名の話者における自然な会話においては、話者同士が交互に発話を行うことでコミュ

ニケーションを図る為、実施例１で説明した通り、一方の話者の発話量が多い時間帯は、他方の話者の発話量が少ないことが想定される。ここで、３名の話者が存在し、３名の話者同士が交互に発話を行う場合、１名が発話している場合、残りの２名の発話量が少ないことが想定される。この場合において、実質的に３名の音声の信号強度の相関係数を算出することで、より高い精度で会話しているユーザの組み合せを判定することが出来る。例えば、図１７において、第２信号強度と第３信号強度が０の時間帯（重複無音区間に相当）が存在する。この時間帯に第１信号強度が０より大きければ、参照相関係数は、第２信号強度と第３信号強度の相関係数未満になる。また、第２信号強度と第３信号強度の重複無音区間が第５閾値以上（例えば、重複無音区間＝１０秒）の場合（換言すると、第１信号強度が０より大きいことが想定される場合）のみに、参照相関係数を算出する処理を実施しても良い。当該処理により算出部４の算出処理の負荷を軽減することができる。また、重複無音区間において、信号強度が大きいフレームを最も多く含む音声の信号強度を、優先的に参照相関係数を算出する信号強度として選択しても良い。

【００６７】

実施例６における音声処理装置１によれば、複数の話者の音声が個別に録音された音声データから、会話が行われている話者に対応した音声データの組み合せを、高い精度で判定することが可能となる。なお、実施例６における音声処理装置１は、実施例１ないし実施例５に記載した音声処理を任意に組み合せることができる。

【００６８】

（実施例７）

図１の検出部３は、実施例1ないし実施例６の検出処理に加え、第１入力信号に含まれる第２信号強度または、第２入力信号に含まれる第１信号強度を更に検出することができる。また、算出部４は、実施例1ないし実施例６の算出処理に加え、第１入力信号に含まれる第２信号強度と第２入力信号に含まれる第２信号強度、または、第２信号に含まれる１信号強度と第１信号に含まれる第１信号強度の時系列に対する第２相関係数を更に算出することができる。更に、判定部５は、実施例１ないし実施例６の判定処理に加え、第２相関係数に基づいて、第１音声と第２音声が会話している状態か否かを判定することができる。以下、実施例７に係る音声処置装置１の音声処理の詳細について説明する。

【００６９】

例えば、第１ユーザと第２ユーザが会議や雑談などの対面での会話している場合、双方の距離間隔は比較的近接している為、第１ユーザが装着しているマイクロフォンには、第１ユーザの第１音声のみならず、第２ユーザの第２音声も入力される場合もある。同様に、第２ユーザが装着しているマイクロフォンには、第２ユーザの第２音声のみならず、第１ユーザの第１音声も入力される場合もある。この為、実施例７においては、検出部３が取得部２から受ける第１信号には第２信号も含まれているものとし、第２信号には第１信号も含まれているものとして説明する。なお、説明の便宜上、第１信号に含まれる第２信号の区間、または、第２信号に含まれる第１信号の区間を回り込み区間と称するものとする。

【００７０】

検出部３は、第１信号に含まれる第２信号を、第１信号から分離する。検出部３は、具体的には、第１音声の第１信号強度の一例となる第１パワーＰ１（ｔ）と、第２音声の第２信号強度の一例となる第２パワーＰ２（ｔ）について、所定の時間間隔（例えば、1秒）の範囲毎に相関係数$corr_n$を算出し、相関性が高いと判定される場合に、当該範囲を回り込み区間と判定することができる。

（数１５）

$$corr_n = \frac{\sum_{i=Tn-1}^{Tn}\left(P1(i)-\overline{P1(i)}\right)\left(P2(i)-\overline{P2(i)}\right)}{\sqrt{\sum_{i=Tn-1}^{Tn}\left(P1(i)-\overline{P1(i)}\right)^2}\sqrt{\sum_{i=Tn-1}^{Tn}\left(P2(i)-\overline{P2(i)}\right)^2}}$$

if $corr_n > TH\_SNEAK$：回り込み区間（$\overline{P1(i)}$と$\overline{P2(i)}$を比較し、小さい方が回り込み区間）
else　　：回り込み区間でない

ここで、上述の（数１５）において、Ｔｎは相関算出範囲のフレーム長（例えば、1フレーム＝２０ｍｓｅｃ）を示し、例えば、１ｓｅｃの場合はＴ＿ｎ＝ｎ＊５０とする。または、回り込み区間の判定閾値となるＴＨ＿ＳＮＥＡＫは、例えば０．９５とすれば良い。また、検出部３は、同様に上述の（数１５）を用いて、第２信号に含まれる第１信号を、次式に基づいて第２信号から分離することができる。検出部３は、第１信号に含まれており、当該第１信号から分離した第２信号、または、第２信号に含まれており、当該第２信号から分離した第１信号を算出部４に出力する。なお、説明の便宜上、第１信号に含まれており、当該第１信号から分離した第２信号を「第２分離信号」と称するものとする。また、第２信号に含まれており、当該第２信号から分離した第１信号を「第１分離信号」と称するものとする。検出部３は、第１分離信号と第２分離信号を算出部４に出力する。

【００７１】

算出部４は、第１分離信号と第２分離信号を検出部３から受け取る。算出部４は、第１分離信号の信号強度と第１信号強度、または、第２分離信号の信号強度と第２信号強度の信号強度の時系列に対する第２相関係数を、上述の（数４）を用いて算出する。算出部４は、算出した第２相関係数を判定部５に出力する。ここで、算出部４が第２相関係数を算出する技術的意義について説明する。例えば、第１分離信号の信号強度と第１信号強度の時系列に対する第２相関係数を考えると、第１分離信号と第１信号は、入力されるマイクロフォンが異なる為、信号強度自体は異なるが、第１ユーザから発せられたものである。この為、第１信号がある程度の強度を有する場合、第１分離信号もある程度の強度を有することになる。この為、例えば、第１分離信号の信号強度と第１信号強度の時系列に対する第２相関係数は正の相関を有することになる。この為、算出部４が、正の相関の強さを示す第２相関係数を算出することで、後述する判定部５が、当該第２相関係数に基づいて、第１音声と第２音声が会話している状態か否かを判定することができる。

【００７２】

判定部５は、第２相関係数を算出部４から受け取る。判定部５は、第２相関係数に基づいて、第１音声と第２音声が会話している状態か否かを判定することができる。例えば、判定部５は、第２相関係数が正であり、当該第２相関係数が所定の第６閾値（例えば第６閾値＝＋０．４）以上の場合に、第１音声と第２音声が会話している状態と判定することができる。また、判定部５は、第２相関係数が第６閾値以上の場合かつ、相関係数が第１未満の場合に第１音声と第２音声が会話している状態と判定することができる。

【００７３】

実施例７における音声処理装置１によれば、複数の話者の音声が個別に録音された音声データから、会話が行われている話者に対応した音声データの組み合せを、第２相関係数に基づいて判定することで、高い精度で判定することが可能となる。なお、実施例６における音声処理装置１は、実施例１ないし実施例６に記載した音声処理を任意に組み合せることができる。

【００７４】

（実施例８）

図１８は、一つの実施形態による音声処理装置１として機能するコンピュータのハードウェア構成図である。図１８に示す通り、音声処理装置１は、コンピュータ１００、およびコンピュータ１００に接続する入出力装置（周辺機器）を含んで構成される。

【００７５】

コンピュータ１００は、プロセッサ１０１によって装置全体が制御されている。プロセッサ１０１には、バス１０９を介してＲＡＭ（Ｒａｎｄｏｍ　Ａｃｃｅｓｓ　Ｍｅｍｏｒｙ）１０２と複数の周辺機器が接続されている。なお、プロセッサ１０１は、マルチプロセッサであってもよい。また、プロセッサ１０１は、例えば、ＣＰＵ、ＭＰＵ（Ｍｉｃｒ

ｏ Ｐｒｏｃｅｓｓｉｎｇ Ｕｎｉｔ）、ＤＳＰ（Ｄｉｇｉｔａｌ Ｓｉｇｎａｌ Ｐｒｏｃｅｓｓｏｒ）、ＡＳＩＣ（Ａｐｐｌｉｃａｔｉｏｎ Ｓｐｅｃｉｆｉｃ Ｉｎｔｅｇｒａｔｅｄ Ｃｉｒｃｕｉｔ）、またはＰＬＤ（Ｐｒｏｇｒａｍｍａｂｌｅ Ｌｏｇｉｃ Ｄｅｖｉｃｅ）である。更に、プロセッサ１０１は、ＣＰＵ、ＭＰＵ、ＤＳＰ、ＡＳＩＣ、ＰＬＤのうちの２以上の要素の組み合せであってもよい。なお、例えば、プロセッサ１０１は、図１に記載の取得部２、検出部３、算出部４、判定部５等の機能ブロックの処理を実行することが出来る。例えば、プロセッサ１０１は、コンピュータプログラムにより実現される機能モジュールとして実現された図１の取得部２、検出部３、算出部４、判定部５の処理を実施することができる。

【００７６】

ＲＡＭ１０２は、コンピュータ１００の主記憶装置として使用される。ＲＡＭ１０２には、プロセッサ１０１に実行させるＯＳ（Ｏｐｅｒａｔｉｎｇ Ｓｙｓｔｅｍ）のプログラムやアプリケーションプログラムの少なくとも一部が一時的に格納される。例えば、ＲＡＭ１０２には、図１の取得部２、検出部３、算出部４、判定部５の処理を実現する機能モジュールのプログラムの一部が一時的に格納される。また、ＲＡＭ１０２には、プロセッサ１０１による処理に必要な各種データ（例えば、上述の第１閾値ないし第６閾値や、算出部４が算出する相関係数など）が格納される。バス１０９に接続されている周辺機器としては、ＨＤＤ（Ｈａｒｄ Ｄｉｓｋ Ｄｒｉｖｅ）１０３、グラフィック処理装置１０４、入力インタフェース１０５、光学ドライブ装置１０６、機器接続インタフェース１０７およびネットワークインタフェース１０８がある。

【００７７】

ＨＤＤ１０３は、内蔵したディスクに対して、磁気的にデータの書き込みおよび読み出しを行う。ＨＤＤ１０３は、例えば、コンピュータ１００の補助記憶装置として使用される。ＨＤＤ１０３には、ＯＳのプログラム、アプリケーションプログラム、および各種データが格納される。なお、補助記憶装置としては、フラッシュメモリなどの半導体記憶装置を使用することも出来る。なお、ＨＤＤ１０３に、プロセッサ１０１による処理に必要な各種データ（例えば、上述の第１閾値ないし第６閾値や、図１の算出部４が算出する相関係数など）が格納されても良い。

【００７８】

グラフィック処理装置１０４には、モニタ１１０が接続されている。グラフィック処理装置１０４は、プロセッサ１０１からの命令にしたがって、各種画像をモニタ１１０の画面に表示させる。モニタ１１０としては、ＣＲＴ（Ｃａｔｈｏｄｅ Ｒａｙ Ｔｕｂｅ）を用いた表示装置や液晶表示装置などがある。なお、モニタ１１０は、例えば、図１の判定部５の判定結果を出力する外部装置に相当する。

【００７９】

入力インタフェース１０５には、キーボード１１１とマウス１１２とが接続されている。入力インタフェース１０５は、キーボード１１１やマウス１１２から送られてくる信号をプロセッサ１０１に送信する。なお、マウス１１２は、ポインティングデバイスの一例であり、他のポインティングデバイスを使用することもできる。他のポインティングデバイスとしては、タッチパネル、タブレット、タッチパッド、トラックボールなどがある。キーボード１１１やマウス１１２は、例えば、音声処理装置１を使用するユーザが、音声処理の開始や終了を指示する場合に使用されれば良い。

【００８０】

光学ドライブ装置１０６は、レーザ光などを利用して、光ディスク１１３に記録されたデータの読み取りを行う。光ディスク１１３は、光の反射によって読み取り可能なようにデータが記録された可搬型の記録媒体である。光ディスク１１３には、ＤＶＤ（Ｄｉｇｉｔａｌ Ｖｅｒｓａｔｉｌｅ Ｄｉｓｃ）、ＤＶＤ－ＲＡＭ、ＣＤ－ＲＯＭ（Ｃｏｍｐａｃｔ Ｄｉｓｃ Ｒｅａｄ Ｏｎｌｙ Ｍｅｍｏｒｙ）、ＣＤ－Ｒ（Ｒｅｃｏｒｄａｂｌｅ）／ＲＷ（ＲｅＷｒｉｔａｂｌｅ）などがある。可搬型の記録媒体となる光ディスク１１３に格納されたプログラムは光学ドライブ装置１０６を介して音声処理装置１にインス

トールされる。インストールされた所定のプログラムは、音声処理装置１より実行可能となる。

【００８１】

機器接続インタフェース１０７は、コンピュータ１００に周辺機器を接続するための通信インタフェースである。例えば、機器接続インタフェース１０７には、メモリ装置１１４やメモリリーダライタ１１５を接続することが出来る。メモリ装置１１４は、機器接続インタフェース１０７との通信機能を搭載した記録媒体である。メモリリーダライタ１１５は、メモリカード１１６へのデータの書き込み、またはメモリカード１１６からのデータの読み出しを行う装置である。メモリカード１１６は、カード型の記録媒体である。また、機器接続インタフェース１０７には、マイクロフォン１１８を（有線または無線通信状態で）接続することができる。なお、マイクロフォン１１８は、複数接続されており、例えば、第１ユーザの第１音声や第２ユーザの第２音声が入力される。なお、第１音声や第２音声は、例えば、機器接続インタフェース１０７を介して、プロセッサ１０１に出力されることで、プロセッサ１０１は、コンピュータプログラムにより実現される機能モジュールとして実現された図１の取得部２の処理を実行することができる。

【００８２】

ネットワークインタフェース１０８は、ネットワーク１１７に接続されている。ネットワークインタフェース１０８は、ネットワーク１１７を介して、他のコンピュータまたは通信機器との間でデータの送受信を行う。なお、ネットワークインタフェース１０８は、第１ユーザの第１音声や第２ユーザの第２音声を、ネットワーク１１７を介して他のコンピュータまたは通信機器から受け取っても良い。この場合、なお、第１音声や第２音声は、例えば、ネットワークインタフェース１０８を介して、プロセッサ１０１に出力されることで、プロセッサ１０１は、コンピュータプログラムにより実現される機能モジュールとして実現された図１の取得部２の処理を実行することができる。

【００８３】

コンピュータ１００は、たとえば、コンピュータ読み取り可能な記録媒体に記録されたプログラムを実行することにより、上述した音声処理機能を実現する。コンピュータ１００に実行させる処理内容を記述したプログラムは、様々な記録媒体に記録しておくことが出来る。上記プログラムは、１つのまたは複数の機能モジュールから構成することが出来る。例えば、図１に記載の取得部２、検出部３、算出部４、判定部５等の処理を実現させた機能モジュールからプログラムを構成することが出来る。なお、コンピュータ１００に実行させるプログラムをＨＤＤ１０３に格納しておくことができる。プロセッサ１０１は、ＨＤＤ１０３内のプログラムの少なくとも一部をＲＡＭ１０２にロードし、プログラムを実行する。また、コンピュータ１００に実行させるプログラムを、光ディスク１１３、メモリ装置１１４、メモリカード１１６などの可搬型記録媒体に記録しておくことも出来る。可搬型記録媒体に格納されたプログラムは、例えば、プロセッサ１０１からの制御により、ＨＤＤ１０３にインストールされた後、実行可能となる。またプロセッサ１０１が、可搬型記録媒体から直接プログラムを読み出して実行することも出来る。

【００８４】

以上に図示した各装置の各構成要素は、必ずしも物理的に図示の如く構成されていることを要しない。すなわち、各装置の分散・統合の具体的形態は図示のものに限られず、その全部または一部を、各種の負荷や使用状況などに応じて、任意の単位で機能的または物理的に分散・統合して構成することができる。また、上記の実施例で説明した各種の処理は、予め用意されたプログラムをパーソナルコンピュータやワークステーションなどのコンピュータで実行することによって実現することができる。

【００８５】

以上、説明した実施形態に関し、更に以下の付記を開示する。

（付記１）

第１音声が含まれる第１入力信号と、第２音声が含まれる第２入力信号を取得する取得部と、

前記第１入力信号の第１信号強度と、前記第２入力信号の第２信号強度を検出する検出部と、
前記第１信号強度の時系列と前記第２信号強度の時系列との相関係数を算出する算出部と、
前記相関係数に基づいて、前記第１音声と前記第２音声が会話している状態か否かを判定する判定部
を備えることを特徴とする音声処理装置。
（付記２）
前記判定部は、前記相関係数が負であり、前記相関係数が所定の第１閾値未満の場合に、前記第１音声と前記第２音声が会話している状態と判定することを特徴とする付記１記載の音声処理装置。
（付記３）
前記第１信号強度は前記第１音声のパワーまたは信号対雑音比であり、前記第２信号強度は前記第２音声のパワーまたは信号対雑音比であることを特徴とする付記１または付記２に記載の音声処理装置。
（付記４）
前記検出部は、
前記第１入力信号に含まれる第１発話区間を前記第１信号強度に基づいて検出し、
前記第２入力信号に含まれる第２発話区間を前記第２信号強度に基づいて検出し、
前記第１発話区間と前記第２発話区間が重複する重複発話区間を検出し、
前記算出部は、
前記重複発話区間が所定の第２閾値未満の前記第１信号強度と前記第２信号強度を、前記相関係数の算出に用いないことを特徴とする付記１ないし付記３の何れか一つに記載の音声処理装置。
（付記５）
前記検出部は、
前記第１入力信号に含まれる第１無音区間を前記第１信号強度に基づいて検出し、
前記第２入力信号に含まれる第２無音区間を前記第２信号強度に基づいて検出し、
前記第１無音区間と前記第２無音区間が重複する重複無音区間を検出し、
前記算出部は、
前記重複無音区間が所定の第３閾値未満の前記第１信号強度と前記第２信号強度を、前記相関係数の算出に用いないことを特徴とする付記１ないし付記４の何れか一つに記載の音声処理装置。
（付記６）
前記算出部は、前記第１信号強度の第１位相または、前記第２信号強度の第２位相を所定の範囲で変化させて複数の前記相関係数を算出し、
前記判定部は、複数の前記相関係数のうち最小値となる前記相関係数に基づいて、前記第１音声と前記第２音声が会話している状態か否かを判定することを特徴とする付記１ないし付記５の何れか一つに記載の音声処理装置。
（付記７）
前記検出部は、
前記第１信号強度の時系列と前記第２信号強度の時系列との大小関係を比較し、
前記大小関係が反転する反転回数が所定の第４閾値以上となる相関係数算出区間を検出し、
前記算出部は、
前記相関係数算出区間における前記第１信号強度の時系列と前記第２信号強度の時系列との前記相関係数を算出することを特徴とする付記１ないし付記６の何れか一つに記載の音声処理装置。
（付記８）
前記取得部は、第３音声が含まれる第３入力信号を更に取得し、

前記検出部は、前記第３入力信号の第３信号強度を更に検出し、
前記算出部は、前記第１信号強度の時系列、前記第２信号強度の時系列または前記第３信号強度の時系列のうち、２つの信号強度の組み合わせに対する複数の前記相関係数を算出する算出部と、
前記判定部は、複数の前記相関係数のうち、前記相関係数が最小値となる２つの前記信号強度の組み合せに基づいて、前記第１音声、前記第２音声または前記第３音声から、会話している音声の組み合わせを判定することを特徴とする付記１ないし付記７の何れか一つに記載の音声処理装置。
（付記９）
前記算出部は、前記最小値となる前記相関係数の算出に用いた２つの前記信号強度を加算した加算信号強度を算出し、
前記加算信号強度の時系列と、前記最小値となる前記相関係数の算出に用いた２つの前記信号強度以外の１つの信号強度の時系列との参照相関係数を算出し、
前記判定部は、前記最小値となる前記相関係数が前記参照相関係数を下回る場合、または、前記第１閾値未満の場合に、前記参照相関係数の算出に用いた３つの前記信号強度に基づいて、会話している音声の組み合わせを判定することを特徴とする付記８記載の音声処理装置。
（付記１０）
前記算出部は、前記最小値である前記相関係数の算出に用いた２つの前記信号強度の前記重複無音区間が第５閾値以上の場合に、前記参照相関係数を算出することを特徴とする付記９記載の音声処理装置。
（付記１１）
前記検出部は、
　前記第１入力信号に含まれる前記第２信号強度または、前記第２信号に含まれる前記第１信号強度を更に検出し、
前記算出部は、
　前記第１入力信号に含まれる前記第２信号強度の時系列と前記第２入力信号に含まれる前記第２信号強度の時系列との第２相関係数、または、
　前記第２入力信号に含まれる前記第１信号強度の時系列と前記第１入力信号に含まれる前記第１信号強度の時系列との第３相関係数を更に算出し、
前記判定部は、前記第２相関係数または前記第３相関係数に基づいて、前記第１音声と前記第２音声が会話している状態か否かを判定することを特徴とする付記１ないし付記１０の何れか一つに記載の音声処理装置。
（付記１２）
前記判定部は、前記相関係数が正であり、前記第２相関係数が所定の第６閾値以上の場合に、前記第１音声と前記第２音声が前記会話している状態と判定することを特徴とする付記１１記載の音声処理装置。
（付記１３）
第１音声が含まれる第１入力信号と、第２音声が含まれる第２入力信号を取得し、
前記第１入力信号の第１信号強度と、前記第２入力信号の第２信号強度を検出し、
前記第１信号強度の時系列と前記第２信号強度の時系列との相関係数を算出し、
前記相関係数に基づいて、前記第１音声と前記第２音声が会話している状態か否かを判定することを含む音声処理方法。
（付記１４）
前記判定することは、前記相関係数が負であり、前記相関係数が所定の第１閾値未満の場合に、前記第１音声と前記第２音声が前記会話している状態と判定することを特徴とする付記１３記載の音声処理方法。
（付記１５）
前記第１信号強度または前記第２信号強度は、前記第１音声または前記第２音声のパワーまたは信号対雑音比であることを特徴とする付記１３または付記１４に記載の音声処理

方法。
（付記１６）
前記検出することは、
前記第１入力信号に含まれる第１発話区間を前記第１信号強度に基づいて検出し、
前記第２入力信号に含まれる第２発話区間を前記第２信号強度に基づいて検出し、
前記第１発話区間と前記第２発話区間が重複する重複発話区間を検出し、
前記算出部は、
前記重複発話区間が所定の第２閾値未満の前記第１信号強度と前記第２信号強度を、前記相関係数の算出に用いないことを特徴とする付記１３ないし付記１５の何れか一つに記載の音声処理方法。
（付記１７）
前記検出することは、
前記第１入力信号に含まれる第１無音区間を前記第１信号強度に基づいて検出し、
前記第２入力信号に含まれる第２無音区間を前記第２信号強度に基づいて検出し、
前記第１無音区間と前記第２無音区間が重複する重複無音区間を検出し、
前記算出することは、
前記重複無音区間が所定の第３閾値未満の前記第１信号強度と前記第２信号強度を、前記相関係数の算出に用いないことを特徴とする付記１３ないし付記１６の何れか一つに記載の音声処理方法。
（付記１８）
前記算出することは、前記第１信号強度の第１位相または、前記第２信号強度の第２位相を所定の範囲で変化させて複数の前記相関係数を算出し、
前記判定することは、複数の前記相関係数のうち最小値となる前記相関係数に基づいて、前記第１音声と前記第２音声が会話している状態か否かを判定することを特徴とする付記１３ないし付記１７の何れか一つに記載の音声処理方法。
（付記１９）
前記検出することは、
前記第１信号強度の時系列と前記第２信号強度の時系列との大小関係を比較し、
前記大小関係が反転する反転回数が所定の第４閾値以上となる相関係数算出区間を検出し、
前記算出部は、
前記相関係数算出区間における前記第１信号強度の時系列と前記第２信号強度の時系列との前記相関係数を算出することを特徴とする付記１３ないし付記１８の何れか一つに記載の音声処理方法。
（付記２０）
前記取得することは、第３音声が含まれる第３入力信号を更に取得し、
前記検出することは、前記第３入力信号の第３信号強度を更に検出し、
前記算出することは、前記第１信号強度の時系列、前記第２信号強度の時系列または前記第３信号強度の時系列から、２つの信号強度の組み合わせに対する複数の前記相関係数を算出する算出部と、
前記判定部は、複数の前記相関係数のうち、前記相関係数が最小値となる２つの前記信号強度の組み合わせに基づいて、前記第１音声、前記第２音声または前記第３音声から、会話している音声の組み合わせを判定することを特徴とする付記１３ないし付記１９の何れか一つに記載の音声処理方法。
（付記２１）
前記算出することは、前記最小値となる前記相関係数の算出に用いた２つの前記信号強度を加算した加算信号強度を算出し、
前記加算信号強度の時系列と、前記最小値となる前記相関係数の算出に用いた２つの前記信号強度以外の１つの信号強度の時系列との参照相関係数を算出し、
前記判定することは、前記最小値となる前記相関係数が前記参照相関係数を下回る場合

、または、前記第１閾値未満の場合に、前記参照相関係数の算出に用いた３つの前記信号強度に基づいて、会話している音声の組み合わせを判定することを特徴とする付記２０記載の音声処理方法。

（付記２２）

前記算出することは、前記最小値となる前記相関係数の算出に用いた２つの前記信号強度の前記重複無音区間が第５閾値以上の場合に、前記参照相関係数を算出することを特徴とする付記２１記載の音声処理方法。

（付記２３）

前記検出することは、

前記第１入力信号に含まれる前記２信号強度または、前記第２入力信号に含まれる前記第１信号強度を更に検出し、

前記算出することは、

前記第１入力信号に含まれる前記第２信号強度の時系列と前記第２入力信号に含まれる前記第２信号強度の時系列との第２相関係数、または、

前記第２入力信号に含まれる前記第１信号強度の時系列と前記第１入力信号に含まれる前記第１信号強度の時系列との第３相関係数を更に算出し、

前記判定部は、前記第２相関係数または前記第３相関係数に基づいて、前記第１音声と前記第２音声が会話している状態か否かを判定することを特徴とする付記１３ないし付記２２の何れか一つに記載の音声処理方法。

（付記２４）

前記判定することは、前記相関係数が正であり、前記第２相関係数が所定の第６閾値以上の場合に、前記第１音声と前記第２音声が前記会話している状態と判定することを特徴とする付記２３記載の音声処理方法。

（付記２５）

コンピュータに

第１音声が含まれる第１入力信号と、第２音声が含まれる第２入力信号を取得し、

前記第１入力信号の第１信号強度と、前記第２入力信号の第２信号強度を検出し、

前記第１信号強度の時系列と前記第２信号強度の時系列との相関係数を算出し、

前記相関係数に基づいて、前記第１音声と前記第２音声が会話している状態か否かを判定する

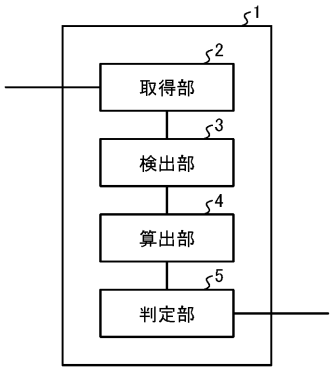
ことを実行させることを特徴とする音声処理プログラム。

【符号の説明】

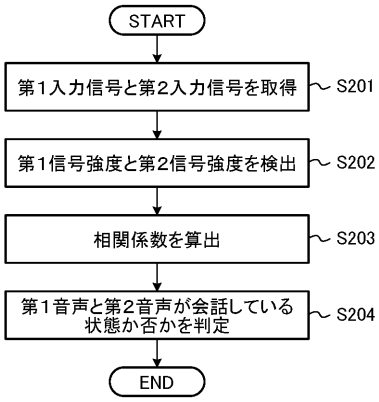
【００８６】

１　音声処理装置
２　取得部
３　検出部
４　算出部
５　判定部

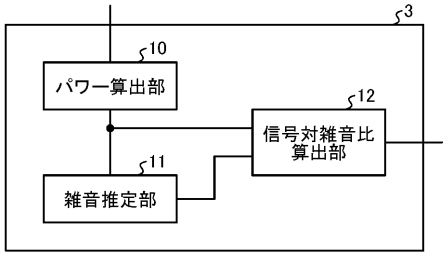
【図１】


1
2
取得部
3
検出部
4
算出部
5
判定部

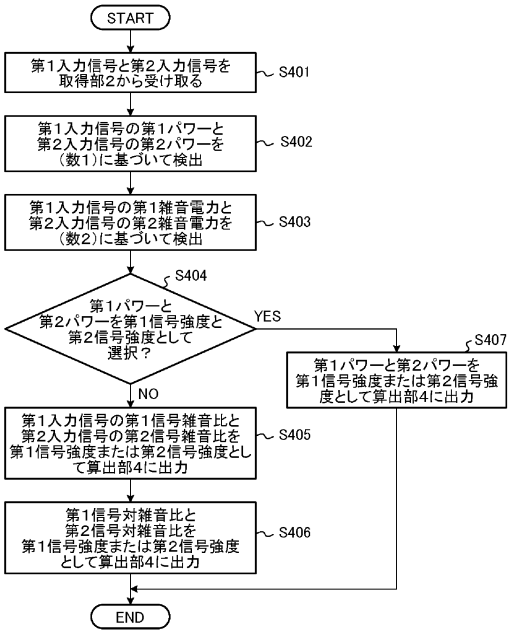
【図２】


START
第１入力信号と第２入力信号を取得
S201
第１信号強度と第２信号強度を検出
S202
相関係数を算出
S203
第１音声と第２音声が会話している状態か否かを判定
S204
END

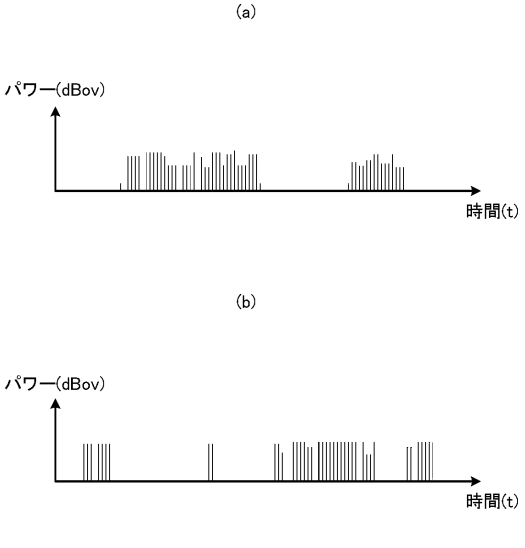
【図３】


3
10
パワー算出部
12
信号対雑音比算出部
11
雑音推定部

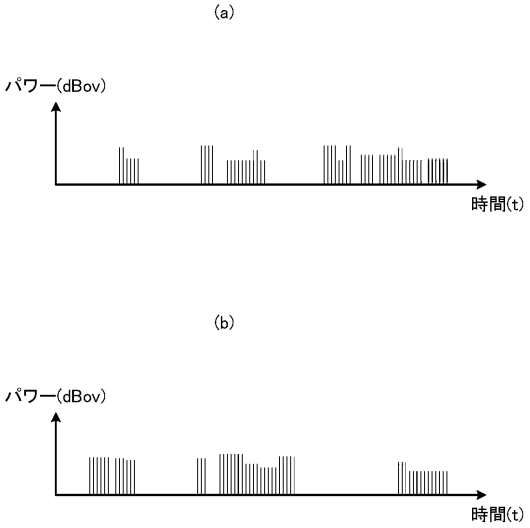
【図４】


START
第１入力信号と第２入力信号を取得部２から受け取る
S401
第１入力信号の第１パワーと第２入力信号の第２パワーを（数１）に基づいて検出
S402
第１入力信号の第１雑音電力と第２入力信号の第２雑音電力を（数２）に基づいて検出
S403
S404
第１パワーと第２パワーを第１信号強度と第２信号強度として選択？
YES
S407
第１パワーと第２パワーを第１信号強度または第２信号強度として算出部４に出力
NO
第１入力信号の第１信号雑音比と第２入力信号の第２信号雑音比を第１信号強度または第２信号強度として算出部４に出力
S405
第１信号対雑音比と第２信号対雑音比を第１信号強度または第２信号強度として算出部４に出力
S406
END

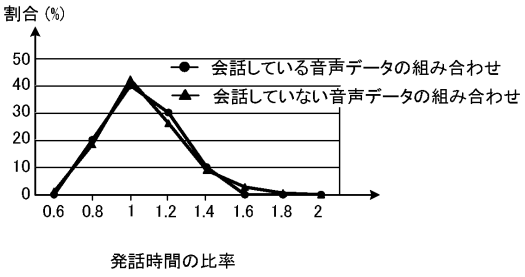
【図５】


(a)
パワー(dBov)
時間(t)
(b)
パワー(dBov)
時間(t)

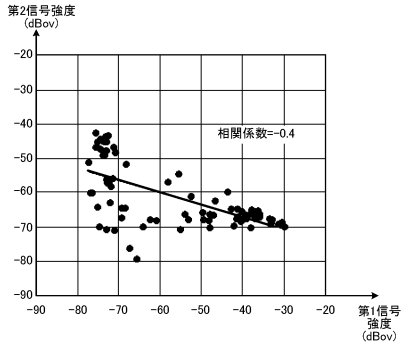
【図６】


(a)
パワー(dBov)
時間(t)
(b)
パワー(dBov)
時間(t)

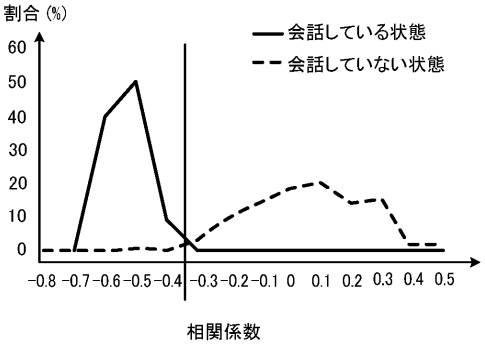
【図７】


割合 (%)
50
40
30
20
10
0
0.6 0.8 1 1.2 1.4 1.6 1.8 2
会話している音声データの組み合わせ
会話していない音声データの組み合わせ
発話時間の比率

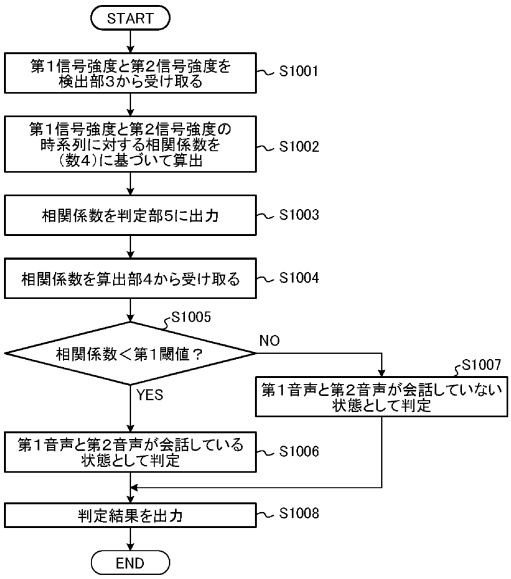
【図８】


第2信号強度
(dBov)
-20
-30
-40
-50
-60
-70
-80
-90
相関係数=-0.4
-90 -80 -70 -60 -50 -40 -30 -20
第1信号
強度
(dBov)

【図９】


割合 (%)
60
50
40
30
20
10
0
会話している状態
会話していない状態
-0.8 -0.7 -0.6 -0.5 -0.4 -0.3 -0.2 -0.1 0 0.1 0.2 0.3 0.4 0.5
相関係数

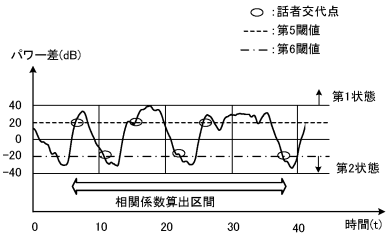
【図１０】



【図１１】

	検出率（%）	正解率（%）
比較例	100	5
実施例	100	91

【図１２】



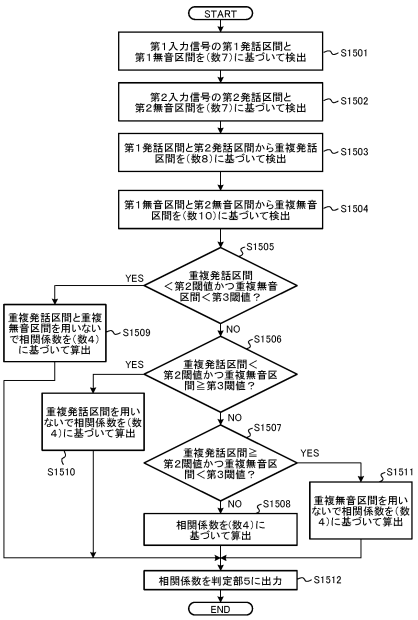
【図１３】

相関係数算出区間に含まれる話者交替回数	会話している状態における相関係数	会話していない状態における相関係数
8＞	−0.65	0.00
8	−0.65	0.00
6	−0.60	0.04
4	−0.56	−0.60
2	−0.36	−0.45

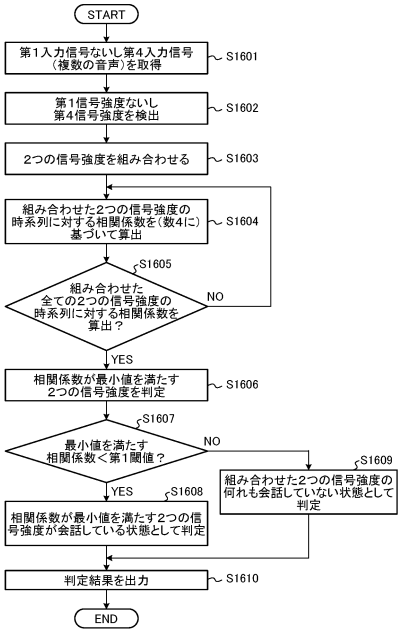
【図１４】

組み合せID	信号強度の組み合せ		相関係数
1	第1信号強度	第2信号強度	−0.54
2	第1信号強度	第3信号強度	0.21
3	第1信号強度	第4信号強度	0.46
4	第2信号強度	第3信号強度	−0.61
5	第2信号強度	第4信号強度	−0.21
6	第3信号強度	第4信号強度	0.63

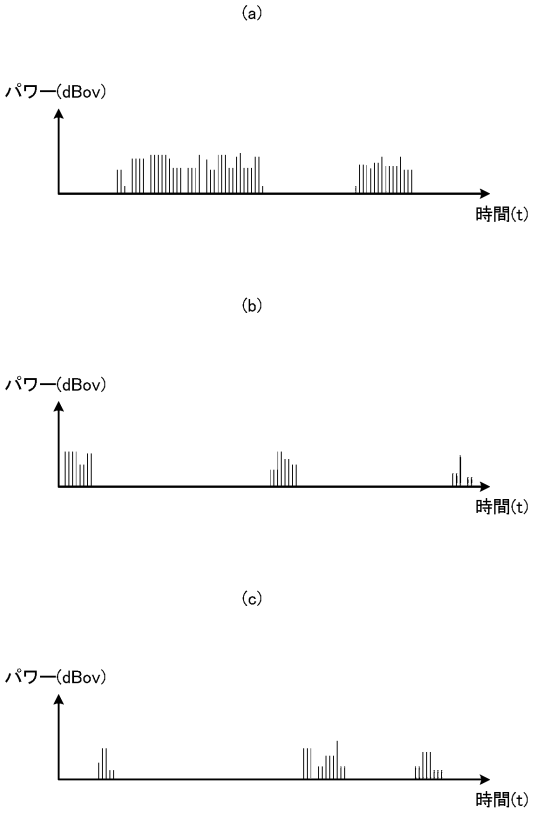
【図１５】



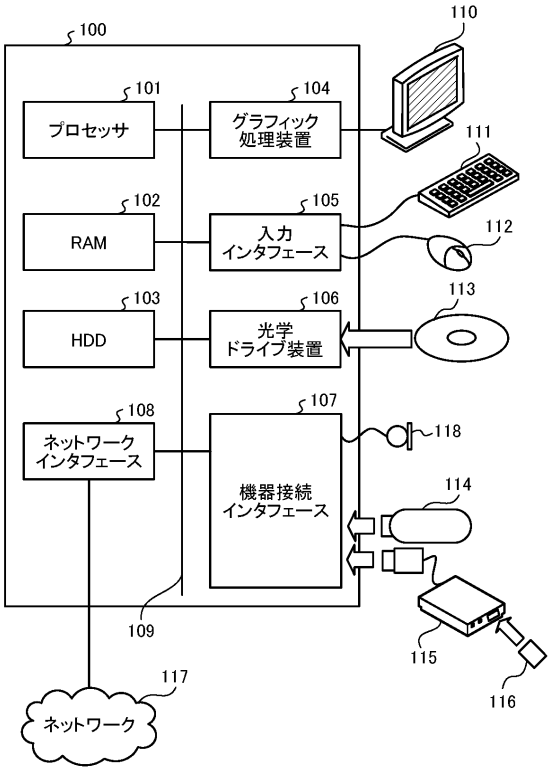
【図１６】


START
第1入力信号ないし第4入力信号（複数の音声）を取得
S1601
第1信号強度ないし第4信号強度を検出
S1602
2つの信号強度を組み合わせる
S1603
組み合わせた2つの信号強度の時系列に対する相関係数を（数4に）基づいて算出
S1604
S1605
組み合わせた全ての2つの信号強度の時系列に対する相関係数を算出？
NO
YES
相関係数が最小値を満たす2つの信号強度を判定
S1606
S1607
最小値を満たす相関係数＜第1閾値？
NO
YES
S1608
相関係数が最小値を満たす2つの信号強度が会話している状態として判定
S1609
組み合わせた2つの信号強度の何れも会話していない状態として判定
判定結果を出力
S1610
END

【図１７】


(a)
パワー(dBov)
時間(t)
(b)
パワー(dBov)
時間(t)
(c)
パワー(dBov)
時間(t)

【図１８】


100
101
プロセッサ
102
RAM
103
HDD
104
グラフィック処理装置
105
入力インタフェース
106
光学ドライブ装置
107
機器接続インタフェース
108
ネットワークインタフェース
109
110
111
112
113
114
115
116
117
ネットワーク
118

フロントページの続き

審査官　山下　剛史

(56)参考文献　特開２０１３－１４０５３４（ＪＰ，Ａ）
特開２００８－２４２３１８（ＪＰ，Ａ）
特開２００５－２０２０３５（ＪＰ，Ａ）
特表２０１２－５０３４００（ＪＰ，Ａ）
特開２００４－１３３４０３（ＪＰ，Ａ）
特開２０１３－１１５６２２（ＪＰ，Ａ）
Aki HARMA, et al., CONVERSATION DETECTION IN AMBIENT TELEPHONY, ICASSP 2009, IEEE, ２００９年　4月, p.4641-4644
川口洋平他，音源分離を利用した話者交替の性質に基づく会話抽出，日本音響学会２０１０年春季研究発表会講演論文集［ＣＤ－ＲＯＭ］，２０１０年　3月，p.869-870

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ１０Ｌ　１３／００－９９／００
ＩＥＥＥ　Ｘｐｌｏｒｅ