

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第6524674号  
(P6524674)

(45) 発行日 令和1年6月5日(2019.6.5)

(24) 登録日 令和1年5月17日(2019.5.17)

(51) Int.Cl.	F I
G 1 O L 25/63 (2013.01)	G 1 O L 25/63
H O 4 M 1/00 (2006.01)	H O 4 M 1/00 V
	H O 4 M 1/00 H

請求項の数 12 (全 20 頁)

(21) 出願番号	特願2015-10342 (P2015-10342)	(73) 特許権者	000005223
(22) 出願日	平成27年1月22日 (2015.1.22)		富士通株式会社
(65) 公開番号	特開2016-133774 (P2016-133774A)		神奈川県川崎市中原区上小田中4丁目1番1号
(43) 公開日	平成28年7月25日 (2016.7.25)	(74) 代理人	100147164
審査請求日	平成29年12月15日 (2017.12.15)		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	塩田 千里
			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

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ユーザの第1音声に含まれる第1発話区間と、第2ユーザの第2音声に含まれる第2発話区間を検出する検出部と、

前記第1発話区間と前記第2発話区間が重複する重複区間を規定する規定部と、

前記重複区間の始点から前記第1発話区間の終点までの第1発話継続区間を算出する算出部と、

前記第1発話継続区間の長さに関する情報に基づいて、前記第1音声に対する印象を評価する評価部

を備えることを特徴とする音声処理装置。

10

【請求項2】

前記規定部は、前記第1発話区間と前記第2発話区間が第1閾値以上の長さで重複する場合に、前記重複区間を規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】

前記規定部は、前記第1発話区間または前記第2発話区間に含まれる母音数を算出し、前記第1発話区間と前記第2発話区間が第2閾値以上の前記母音数で重複する場合に、前記重複区間を規定す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音声と前記第2音声に含まれる複数のフレームの第1信号品質と第2信号品質をそれぞれ検出し、

20

前記規定部は、前記第 1 信号品質と前記第 2 信号品質が第 3 閾値以上であり、かつ、前記第 1 発話区間と前記第 2 発話区間が重複する区間を前記重複区間として規定する請求項 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】

前記検出部は、前記第 1 音声と前記第 2 音声に含まれる複数のフレームの第 1 信号品質と第 2 信号品質をそれぞれ検出し、

前記規定部は、前記第 1 信号品質と前記第 2 信号品質の差が第 4 閾値未満であり、かつ、前記第 1 発話区間と前記第 2 発話区間が重複する区間を前記重複区間として規定する請求項 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 6】

前記第 1 信号品質または前記第 2 信号品質は、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請求項 4 または請求項 5 に記載の音声処理装置。

【請求項 7】

前記規定部は、前記第 1 発話区間と、前記第 2 発話区間のいずれか早い区間の開始時刻から第 5 閾値の時間経過後に前記重複区間を規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】

前記規定部は、前記第 1 発話区間と、前記第 2 発話区間のいずれか遅い区間の終了時刻から前記第 5 閾値の時間経過前に前記重複区間を規定することを特徴とする請求項 1 ないし請求項 7 の何れか一項に記載の音声処理装置。

【請求項 9】

前記第 1 音声または前記第 2 音声は、送話音声または受話音声であることを特徴とする請求項 1 ないし請求項 8 の何れか一項に記載の音声処理装置。

【請求項 10】

前記評価部は、前記第 1 発話継続区間の長さが第 6 閾値以上の場合に、前記印象を悪い印象と評価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】

第 1 ユーザの第 1 音声に含まれる第 1 発話区間と、第 2 ユーザの第 2 音声に含まれる第 2 発話区間を検出し、

前記第 1 発話区間と前記第 2 発話区間が重複する重複区間を規定し、

前記重複区間の始点から前記第 1 発話区間の終点までの第 1 発話継続区間を算出し、

前記第 1 発話継続区間の長さに関する情報に基づいて、前記第 1 音声に対する印象を評価する

ことを含むことを特徴とする音声処理方法。

【請求項 12】

コンピュータに、

第 1 ユーザの第 1 音声に含まれる第 1 発話区間と、第 2 ユーザの第 2 音声に含まれる第 2 発話区間を検出し、

前記第 1 発話区間と前記第 2 発話区間が重複する重複区間を規定し、

前記重複区間の始点から前記第 1 発話区間の終点までの第 1 発話継続区間を算出し、

前記第 1 発話継続区間の長さに関する情報に基づいて、前記第 1 音声に対する印象を評価する

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】

10

20

30

40

50

## 【0002】

近年、情報処理機器の発達に伴い、例えば、携帯端末やパーソナルコンピュータにインストールされる通話アプリケーションを介して会話が行われる場面が増加している。相手と自分が会話する場合、お互いの考えを理解し合いながら対話を進めることで円滑なコミュニケーションが実現できる。この場合、相手の考えを十分に理解するためには、相手の発話を遮らずに発話することが必要である。そのため、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上で、入力音声から相手の発話に対する遮りに関する印象を評価する技術が求められている。例えば、企業内での電話会議等において、従業員ごとの発話の印象を評価することにより、組織内のコミュニケーションの問題抽出に活用することができる。

10

## 【先行技術文献】

## 【特許文献】

## 【0003】

【特許文献1】特開2010-175984号公報

## 【発明の概要】

## 【発明が解決しようとする課題】

## 【0004】

第2ユーザの第1ユーザに対する発話の印象を評価する場合、第2ユーザと第1ユーザの会話を、評価者となる第3ユーザが主観的に評価する手法が用いられているが、評価に要する処理時間や、評価の客観性の確保等を考慮すると、必ずしも好ましい手法ではない。しかしながら、第2ユーザの第1ユーザに対する発話の印象を評価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、第2ユーザの第1ユーザに対する発話の印象を評価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。

20

## 【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ユーザの第1音声に含まれる第1発話区間と、第2ユーザの第2音声に含まれる第2発話区間を検出する検出部と、第1発話区間と第2発話区間が重複する重複区間を規定する規定部を備える。更に、当該音声処理装置は、重複区間の始点から第1発話区間の終点までの第1発話継続区間を算出する算出部と、少なくとも第1発話継続区間の長さに関する情報に基づいて、第1音声に対する印象を評価する評価部を備える。

30

## 【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わせにより実現され、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、第2ユーザの第1ユーザに対する発話の印象を評価することが可能となる。

40

## 【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】検出部による第1発話区間と第1無音区間の検出結果を示す図である。

【図5】(a)は、第1発話継続区間の第1概念図である。(b)は、第1発話継続区間の第2概念図である。

【図6】第1発話継続区間と発話の印象の対応テーブルである

【図7】一つの実施形態による携帯端末装置として機能するハードウェア構成図である。

50

【図 8】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、規定部 4、算出部 5、評価部 6 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】

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 S 201 に対応する。ここで、第 1 音声は、例えば、音声処理装置 1 を使用する第 1 ユーザ（自分と称しても良い）の会話相手となる第 2 ユーザ（相手と称しても良い）に対して発話する音声を指す、第 1 ユーザの送話音声である。また、第 2 音声は、例えば、第 2 ユーザが第 1 ユーザに対して発話する音声を指す、第 1 ユーザの受話音声である。なお、送話音声は第 2 ユーザの第 2 音声あり、受話音声は第 1 ユーザの第 1 音声であっても良い。また、取得部 2 は、例えば、音声処理装置 1 に接続または配置される、図示しないマイクロフォン（上述の外部装置に相当）から送話音声を取得することが可能である。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。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 音声と第 2 音声を取得部 2 から受け取る。検出部 3 は、第 1 音声に含まれる第 1 発話区間（または第 1 有音区間と称しても良い）を示す第 1 呼気区間を検出する。また、検出部 3 は、第 2 音声に含まれる第 2 発話区間（または第 2 有音区間と称しても良い）を示す第 2 呼気区間を検出する。なお、当該処理は、図 2 に示すフローチャートのステップ S 202 に対応する。また、第 1 呼気区間または第 2 呼気区間は、例えば、第 1 ユーザまたは第 2 ユーザが発話中において吸気を行ってから発話を開始し、再び吸気を実施するまでの区間（換言すると第 1 の呼吸と第 2 呼吸の間の区間、または発話を続けている区間）となる。検出部 3 は、例えば、第 1 音声または第 2 音声に含まれる複数のフレームから信号品質（第 1 信号対雑音比または第 2 信号対雑音比と称しても良い）の一例となる信号対雑音比となる平均 SNR を検出し、当該平均 SNR が所定の条件を満たしている区間を第 1 発話区間または第 2 発話区間として検出することが出来る。また、検出部 3 は、第 1 音声または第 2 音声に含まれる、第 1 発話区間または第 2 発話区間の後端に接続する無音区間を示す吸気区間を検出する。検出部 3 は、例えば、上述の当該平均 SNR が所定の条件を満たさない区間を第 1 無音区間または第 2 無音区間（換言すると第 1 吸気区間または第 2 吸気区間）として検出することが出来る。

【0013】

ここで、検出部 3 による第 1 発話区間または第 2 発話区間、ならびに第 1 無音区間または第 2 無音区間の検出処理の詳細について説明する。なお、第 2 発話区間の検出方法は、

第1発話区間の検出方法と同様の手法を用いることができ、第2無音区間の検出方法は、第1無音区間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発話区間と第1無音区間の検出処理の詳細について説明する。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、音量算出部9、雑音推定部10、平均SNR算出部11、区間決定部12を有する。なお、検出部3は、音量算出部9、雑音推定部10、平均SNR算出部11、区間決定部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。

10

【0014】

図3において、第1音声は音量算出部9に入力される。なお、音量算出部9は、図示しない長さMのバッファまたはキャッシュを有しても良い。音量算出部9は、第1音声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10と平均SNR算出部11へ出力する。なお、第1音声に含まれる各フレーム長は、例えば20msである。各フレームの音量 $S(n)$ は、次式の通り、算出することが出来る。

(数1)

$$S(n) = \sum_{t=n*M}^{(n+1)*M-1} c(t)^2$$

20

ここで、 $n$ は、第1音声に含まれる音響フレームの入力開始から各フレームに連続して付されるフレーム番号( $n$ は0以上の整数)、 $M$ は1フレームの時間長、 $t$ は時間、 $c(t)$ は、第1音声の振幅(電力)を示す。

【0015】

雑音推定部10は、各フレームの音量 $S(n)$ を音量算出部9から受け取る。雑音推定部10は、各フレームにおける雑音を推定して、雑音推定結果を平均SNR算出部11へ出力する。ここで、雑音推定部10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1)または、(雑音推定方法2)を用いることが出来る。

30

【0016】

(雑音推定方法1)

雑音推定部10は、フレーム $n$ における雑音の大きさ(電力) $N(n)$ を、フレーム $n$ における音量 $S(n)$ 、前フレーム( $n-1$ )における音量 $S(n-1)$ ならびに、雑音の大きさ $N(n-1)$ に基づいて、次式を用いて推定することが出来る。

(数2)

$$N(n) = \begin{cases} \alpha \cdot N(n-1) + (1-\alpha) \cdot S(n) & , (|S(n-1) - S(n)| < \beta \text{ の場合}) \\ N(n-1) & , (\text{上記以外}) \end{cases}$$

40

ここで、 $\alpha$ 、 $\beta$ は、定数であり、実験的に決定されれば良い。例えば、 $\alpha = 0.9$ 、 $\beta = 2.0$ であれば良い。また、雑音電力の初期値 $N(-1)$ も、実験的に決定されれば良い。上述の(数2)において、フレーム $n$ の音量 $S(n)$ が、1つ前のフレーム $n-1$ の音量 $S(n-1)$ に対して一定値以上変化しない場合には、フレーム $n$ の雑音電力 $N(n)$ が更新される。一方、フレーム $n$ の音量 $S(n)$ が、1つ前のフレーム $n-1$ の音量 $S(n-1)$ に対して一定値以上変化する場合には、1つ前のフレーム $n-1$ の雑音電力 $N(n-1)$ をフレーム $n$ の雑音電力 $N(n)$ とする。なお、雑音電力 $N(n)$ を上述の雑音推定結果と称しても良い。

50

## 【 0 0 1 7 】

( 雑音推定方法 2 )

雑音推定部 1 0 は、雑音の大きさの更新を、次式の ( 数 3 ) を用いて、フレーム  $n$  の音量  $S(n)$  と、1 つ前のフレーム  $n - 1$  の雑音電力  $N(n - 1)$  との比に基づいて実施しても良い。

( 数 3 )

$$N(n) = \begin{cases} \alpha \cdot N(n-1) + (1-\alpha) \cdot S(n) & , (S(n) < \gamma \cdot N(n-1) \text{ の場合}) \\ N(n-1) & , (\text{上記以外}) \end{cases}$$

10

ここで、 $\alpha$  は定数であり、実験的に決定されれば良い。例えば、 $\alpha = 2.0$  であれば良い。また、雑音電力の初期値  $N(-1)$  も、実験的に決定されれば良い。上述の ( 数 3 ) において、フレーム  $n$  の音量  $S(n)$  が、1 つ前のフレーム  $n - 1$  の雑音電力  $N(n - 1)$  に対して一定値  $\gamma$  倍未満である場合には、フレーム  $n$  の雑音電力  $N(n)$  を更新する。一方、フレーム  $n$  の音量  $S(n)$  が、1 つ前のフレーム  $n - 1$  の雑音電力  $N(n - 1)$  に対して一定値  $\gamma$  倍以上である場合には、1 つ前のフレーム  $n - 1$  の雑音電力  $N(n - 1)$  をフレーム  $n$  の雑音電力  $N(n)$  とする。

## 【 0 0 1 8 】

図 3 において、平均 SNR 算出部 1 1 は、音量算出部 9 から各フレームの音量  $S(n)$  を受け取り、雑音推定部 1 0 から雑音推定結果となる各フレームの雑音電力  $N(n)$  を受け取る。なお、平均 SNR 算出部 1 1 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  $L$  フレーム分の音量  $S(n)$ 、雑音電力  $N(n)$  を保持する。平均 SNR 算出部 1 1 は、次式を用いて、分析対象時間 ( フレーム ) 内の平均 SNR を算出し、当該平均 SNR を区間決定部 1 2 へ出力する。

20

( 数 4 )

$$SNR(n) = \frac{1}{L} \sum_{i=0}^{L-1} \frac{S(n-i)}{N(n-i)}$$

30

ここで、 $L$  は促音の一般的な長さよりも大きな値に規定すれば良く、例えば、 $0.5 \text{ sec}$  に相当するフレーム数を規定すれば良い。

## 【 0 0 1 9 】

区間決定部 1 2 は、平均 SNR を平均 SNR 算出部 1 1 から受け取る。区間決定部 1 2 は、次式に基づいて、フレーム単位で当該フレームが発話区間か無音区間 ( 非発話区間 ) かを判定し、判定結果  $v_1(t)$  を出力する。

( 数 5 )

$$\begin{aligned} v_1(t) &= 1 \quad (\text{発話区間}) \\ v_1(t) &= 0 \quad (\text{無音区間}) \end{aligned}$$

40

上述の ( 数 5 ) において、 $t$  はフレーム番号を示す。なお、1 フレームは、例えば、 $20 \text{ ms}$  である。また、上述の ( 数 5 ) 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  $v_1(t) = 1$  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  $v_1(t) = 0$  と代入されることを意味する。区間決定部 1 2 は、 $v_1(t) = 1$  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、区間決定部 1 2 は、第 2 音声に関する判定結果  $v_2(t)$  を  $v_1(t)$  と同様の手法を用いて算出し、第 2 発話区間を出力する。

## 【 0 0 2 0 】

50

図4は、検出部3による第1発話区間と第1無音区間の検出結果を示す図である。図4の横軸は時間を示し、縦軸は第1音声の音量（振幅）を示している。図4に示される通り、各第1発話区間の後端に接続する区間が第1無音区間として検出される。また、図4に示される通り、実施例1に開示する検出部3による第1発話区間の検出においては、周囲雑音に合わせて雑音を学習し、SNRに基づいて発話区間を判別している。この為、周囲雑音による第1発話区間の誤検出を防ぐことができる。また、平均SNRを複数フレームから求めることで、第1発話区間で瞬間的に無音となる時間があっても、連続した発話区間として抽出することができる利点を有している。なお、検出部3は、国際公開第2009/145192号パンフレットに記載の方法を用いることも可能である。なお、上述の通り、第2発話区間の検出方法は、第1発話区間の検出方法と同様の手法を用いること

10

【0021】

図1において、規定部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、規定部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。規定部4は、検出部3が検出した第1発話区間と第2発話区間を検出部3から受け取る。規定部4は、第1発話区間と第2発話区間が重複する重複区間を規定する。なお、当該処理は、図2に示すフローチャートのステップS203に対応する。当該重複区間は、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、規定部4は、具体的には、次式に基づいて重複区間 $L(t)$ を規定することが出来る。

20

(数6)

$$\begin{array}{ll} \text{if } \{v_1(t)=0\} \vee \{v_2(t)=0\} & L(t)=0 \\ \text{else} & L(t)=L(t-1)+1 \end{array}$$

上述の(数6)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（非発話区間）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。規定部4は、規定した重複区間を算出部5に出力する。

30

【0022】

図1において、算出部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、算出部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。算出部5は、規定部4が規定した重複区間を規定部4から受け取る。また、算出部5は、第1発話区間と第2発話区間を検出部3から受け取る。算出部5は、重複区間の始点から第1発話区間の終点までの区間となる第1発話継続区間を算出する。なお、当該処理は、図2に示すフローチャートのステップS204に対応する。具体的には、算出部5は以下の算出処理を実行する。

40

【0023】

算出部5は、重複区間 $L(t)$ と任意の第1閾値 $TH\_L$ との比較結果に基づき重複区間の始点時刻 $T_s$ （重複発生区間時刻 $T_s$ と称しても良い）を出力する。算出部5は、例えば、次式に従い、重複区間（の長さ）が第1閾値以上となる場合に重複区間が発生したと判定し、重複区間の始点時刻 $T_s$ を出力する。算出部5は、第1閾値未満である場合は、重複区間が発生していないと判定し、重複発生区間時刻 $T_s$ を次式に従ってリセットする。ここで、第1閾値は、一般的な相槌の長さに相当する値であれば良い。相槌は、相手の発話に同意の旨等を示すものである為、会話を遮るものとして取扱う必要はない。なお

50

、第1閾値  $TH\_L$  は、例えば1秒(50フレームに相当)とすることが出来る。

(数7)

$$\begin{aligned} \text{if } \{L(t) \geq TH\_L\} & \quad T_s = t - TH\_L \\ \text{else} & \quad T_s = -1 \end{aligned}$$

上述の(数7)は、 $L(t) \geq TH\_L$ の条件を満たす場合は、重複区間が発生して、算出部5が重複区間の始点  $T_s$  を算出することを示し、 $L(t) < TH\_L$ の条件を満たさない場合は、重複区間が発生していないことを示す。

10

【0024】

算出部7は、上述の(数7)において、 $L(t) \geq TH\_L$ の条件を満たす場合は、上述の(数5)を用いて算出した  $v_1(t)$  に基づいて、第1発話区間の終点となる時刻  $T_e$  を次式に基づいて算出する。

(数8)

$$\begin{aligned} \text{if } \{v_1(t-1)=1\} \wedge \{v_1(t)=0\} & \quad T_e = t-1 \\ \text{else} & \quad T_e = -1 \end{aligned}$$

20

上述の(数8)は、直前フレーム( $t-1$ )が、発話区間と判定され、かつ、現フレーム( $t$ )が、無音区間(非発話区間)と判定された場合に発話終了と判定し、直前フレーム( $t-1$ )を第1発話区間の終点となる時刻  $T_e$  として算出することを意味する。また、それ以外の場合は、発話区間が継続中または非発話区間であるため、発話終了時刻  $T_e$  をリセットすることを意味する。

【0025】

算出部5は、第1発話区間の終点の時刻  $T_e$  を算出した場合、重複区間の始点時刻  $T_s$  と第1発話区間の終点の時刻  $T_e$  から第1発話継続区間  $CL$  を次式に基づいて算出する。

(数9)

$$CL = T_e - T_s$$

30

【0026】

図5(a)は、第1発話継続区間の第1概念図である。図5(b)は、第1発話継続区間の第2概念図である。図5(a)、図5(b)においては、第1ユーザが、第2ユーザよりも先に発話を開始している。また、実線で囲まれている領域は、第1ユーザまたは第2ユーザが実際に意図した内容を「発話している領域」であり、点線で囲まれている領域は、重複区間の発生により、第1ユーザまたは第2ユーザが実際に意図した内容を「発話していない領域」である。図5(a)において、第1ユーザの第1音声(実際に発話している音声)は、例えば、「週末は山に行きましょう。紅葉がとても綺麗ですよ。」である。また、第2ユーザの第2音声(実際に発話している音声)は、「でも天気は」である。この時、第2ユーザは、「悪いので別の日に山に行きましょう)は発話せず、第1ユーザは、第1ユーザの発話中に、第2ユーザが発話を開始したのにも係らず、発話を止めることなく、発話を最後まで続けている状態となる。この場合、詳細は後述するが、第2ユーザの第1ユーザに対する発話の印象は悪いものとなる。

40

【0027】

図5(b)において、第1ユーザの第1音声(実際に発話している音声)は、例えば、「週末は山に行きましょう。」である。また、第2ユーザの第2音声(実際に発話している音声)は、「でも天気が悪いので、別の日に山に行きましょう」である。この時、第1ユーザは、第1ユーザの発話中に、第2ユーザが発話を開始した為、発話を最後まで続けることなく、発話を途中で止めている。この場合、詳細は後述するが、第2ユーザの第1ユーザに対する発話の印象は悪いものとならない。実施例1において、第1ユーザが発話

50

中に発話を第2ユーザに譲った場合には、発話区間が重複する重複区間の始点から、第1ユーザが発話を中断するまでの時間が短い性質を利用することにより、第1ユーザの発話の印象を評価することが可能となる。算出部5は、算出した第1発話継続区間を評価部6に出力する。

【0028】

図1において、評価部6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、評価部6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。評価部6は、算出部5が算出した第1発話継続区間を算出部5から受け取る。評価部6は、少なくとも第1発話継続区間の長さに関する情報に基づいて、第2ユーザの第1音声に対する印象を評価する。なお、当該処理は、図2

10

【0029】

図6は、第1発話継続区間と発話の印象の対応テーブルである。なお、図6の対応テーブル60は、本発明者らの鋭意研究の結果、新たに見出されたものである。また、図6の対応テーブル60は、複数の評価者の主観的評価に基づくものである。なお、評価方法は、第1ユーザが発話し続けている時に、第2ユーザに発話にさせて重複区間を発生させた場合における第1発話継続区間(第1発話継続時間と称しても良い)に基づいて、第2ユーザの第1ユーザに対する発話の印象を実験的に求めた。なお、主観的評価は、「印象が悪くない(発話が実質的に遮られていない)」、「印象がやや悪い(発話が実質的にやや遮られている)」、「印象が悪い(発話が実質的に遮られている)」の3段階に分けて行

20

【0030】

評価部6は、第1発話継続区間と図6のテーブル60に基づいて、評価結果を任意の外部装置に出力する。なお、評価部6は、評価結果を任意のスコアに置き換えて、当該スコアを用いて発話を評価しても良い。また、外部装置は、例えば、評価結果を音声に変換させた上で再生できるスピーカや、評価音声文字情報として表示するディスプレイであれば良い。また、評価部6は、評価結果をスピーカに出力する場合は、公知の任意の音声読み上げ方法を用いれば良い。また、評価部6は、評価結果が「やや悪い」または「非常に悪い」場合に、所定の制御信号を外部装置に発信しても良い。なお、当該処理を行う場合は、図2に示すフローチャートのステップS205と合わせて実施すれば良い。また、制御信号は、例えば、アラーム音であれば良い。また、評価部6は、第1発話継続区間の長さが任意の第6閾値(例えば、第6閾値=3秒)以上の場合は、第1ユーザの発話の印象

30

40

【0031】

ここで、実施例1に対応する比較例について説明する。比較例においては、第1発話継続区間を用いずに、単純に重複区間の長さに基づいて第2ユーザの第1ユーザに対する印象を評価する。図5(a)と図5(b)においては、重複区間の長さは同じであることが理解できる。図5(a)においては、重複区間の発生後、第1ユーザは発話を継続し、第2ユーザは発話を停止している。図5(b)においては、重複区間の発生後、第1ユーザは発話を停止し、第2ユーザは発話を継続している。例えば、図6の対応テーブル60の第1発話継続区間の項目を、重複区間の長さに基づいて置き換えて、第1ユーザの発話の印象を評価した場合、図5(a)に示す発話パターンにおいては、第1ユーザの発話の印象は悪い

50

ものと評価されるので特段の問題は生じない。しかしながら、図5(b)に示す発話パターンにおいては、実施例1の評価結果を考慮すると、第1ユーザの発話の印象は悪くないにも係らず、第1ユーザの発話の印象が悪く評価されるといふ誤検出を招く結果となる。この為、比較例に示す様に、単純に第1ユーザと第2ユーザの発話区間の重複時間に基づいて発話の印象を評価する場合は、発話の印象の評価を正確に評価することは難しい。一方、実施例1は、例えば、第1発話継続区間に基づいて発話の印象を評価している為、発話の印象を正確に評価することが出来ることが理解出来る。実施例1における音声処理装置1によれば、第2ユーザの第1ユーザに対する発話の印象を評価することが可能となる。

#### 【0032】

##### (実施例2)

図1の規定部4は、第1発話区間または第2発話区間に含まれる母音数を算出し、当該第1発話区間と第2発話区間が所定の第2閾値(例えば、第2閾値=5)以上の母音数で重複する場合に、重複区間を規定しても良い。換言すると、例えば、「はい」、「いいえ」、「うんうん」、「本当?」、「そうです」等の一般的に相槌に該当する単語の大半の母音数は、例えば、4である為、規定部4は、第1発話区間または第2発話区間に含まれる母音数が第2閾値未満であれば、単なる相槌であり、会話の妨げにならない為、第1発話区間と第2発話区間の重複区間として取扱わないことが出来る。なお、算出部5は、フォルマント分布に基づく母音数の検出方法として、例えば特開2009-258366号公報に記載の方法を用いることが出来る。

#### 【0033】

また、規定部4は、第1発話区間または第2発話区間を文字列として認識しても良い。規定部4は、第1発話区間または第2発話区間を文字列として認識する方法として、例えば、特開平4-255900号公報に開示される方法を適用することが出来る。また、所定の単語は、算出部5が有する図示しないキャッシュまたはメモリに保存される単語リスト(テーブル)に格納される、相槌に相当する単語である。所定の単語は、例えば、「はい」、「いいえ」、「うんうん」、「本当?」、「そうです」等の一般的に相槌に該当する単語であれば良い。規定部4は、第1発話区間または第2発話区間が当該所定の単語と一致する場合、単なる相槌であり、会話の妨げにならない為、第1発話区間と第2発話区間の重複区間として取扱わないことが出来る。なお、音声処理装置1は、実施例2と実施例1の処理を組み合わせた音声処理を行うことが出来る。実施例2における音声処理装置1によれば、第2ユーザの第1ユーザに対する発話の印象をより高い精度で評価することが可能となる。

#### 【0034】

##### (実施例3)

図1の規定部4は、第1発話区間と、第2発話区間のいずれか早い区間の開始時刻から任意の第5閾値(例えば、第5閾値=30秒)の時間経過後に、重複区間を規定しても良い。会話の開始直後は、重複時間が長くても発話印象に影響を与えない挨拶の発話が多いため、発話の印象を判定する対象区間から除外することで判定精度を改善できる。また、規定部4は、第1発話区間と、第2発話区間のいずれか遅い区間の終了時刻から当該第5閾値の経過前に重複区間を規定しても良い。会話終了直前は、重複時間が長くても発話印象に影響を与えない挨拶の発話が多いため、発話の印象を判定する対象区間から除外することで判定精度を改善できる。なお、音声処理装置1は、実施例1ないし実施例3の処理を任意に組み合わせた音声処理を行うことが出来る。実施例3における音声処理装置1によれば、第2ユーザの第1ユーザに対する発話の印象をより高い精度で評価することが可能となる。

#### 【0035】

##### (実施例4)

図1の検出部3は、第1音声と第2音声に含まれる複数のフレームの第1信号品質と第2信号品質をそれぞれ検出し、規定部4は、第1信号品質と第2信号品質が第3閾値(例

10

20

30

40

50

えば、第3閾値 = - 60 dBov) 以上であり、かつ、第1発話区間と第2発話区間が重複する区間を重複区間として規定しても良い。なお、第1信号品質または第2信号品質は、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。例えば、規定部4は、第1ユーザと第2ユーザが共に所定のパワー(音量)以上で発話している場合に、重複区間を規定しても良い。これは、第1ユーザと第2ユーザの何れか一方が、比較的低いパワーで発話している場合、発話の印象に大きな影響を与えないことも想定される為である。検出部3は、上述の(数10)に加えて、次式に基づいて、第1発話区間または、第2発話区間の平均パワーP1とP2を算出することが出来る。なお、平均パワーP1とP2は、それぞれ、第1信号品質と第2信号品質と称しても良い。

(数10)

$$P1(t) = \frac{1}{TH1} \sum_{t=Ts-TH1}^{Ts} s_1^2(t)$$

$$P2(t) = \frac{1}{TH1} \sum_{t=Ts-TH1}^{Ts} s_2^2(t)$$

10

上述の(数10)において、 $s_1(t)$ と $s_2(t)$ は、それぞれ第1音声と第2音声の振幅を表す。規定部4は、第1発話区間ならびに第2発話区間の平均パワーと、任意の閾値TH\_P(第3閾値と称しても良い)との比較結果に基づき、重複区間の発生を判定し、重複区間の始点時刻 $T_s$ を規定することが出来る。具体的には、第1発話区間の平均パワーと第2発話区間の平均パワーが、任意の閾値TH\_P以上の場合は、第1ユーザと第2ユーザがともに所定以上のパワー(大きな音声)で発話している状況である為、発話区間の重複が発生したと判定し、規定部4は、重複区間の始点時刻 $T_s$ を規定する。それ以外の場合は、第1ユーザまたは第2ユーザの何れか一方が小さな声で発話している状況であるため、重複区間が発生していないものとして判定し、規定部4は、重複区間の始点時刻 $T_s$ をリセットする。なお、任意の閾値TH\_Pは例えば、-60 [dBov]とすれば良い。なお、規定部4は、次式に基づいて重複区間を算出することが出来る。

20

(数11)

$$\text{if } \{L(t) \geq TH\_L\} \wedge \{P1(t) \geq TH\_P\} \wedge \{P2(t) \geq TH\_P\} \quad T_s = t - TH\_L$$

$$\text{else} \quad T_s = -1$$

30

上述の(数11)において、 $T_s = t - TH\_L$ は重複区間の発生が有ることを意味し、 $T_s = -1$ は重複区間の発生が無いことを示す。

【0036】

また、規定部4は、第1発話区間と第2発話区間のパワー差に基づいて重複区間を規定することも可能である。例えば、規定部4は、第1発話区間の平均パワーと、第2発話区間の平均パワーの差と、所定の閾値との比較結果に基づき、重複区間の発生を判定し、重複区間の始点時刻 $T_s$ を出力する。重複区間が任意の閾値より長く、第1ユーザの第1発話区間の平均パワーと第2ユーザの第2発話区間の平均パワーとの差が、任意の閾値TH\_P\_DIFF未満(第4閾値と称しても良い)である場合は、第1ユーザと第2ユーザの両者とも同等程度の音量で発話している状況である為、発話の重複区間が発生したと判定し、規定部4は、重複区間の始点時刻 $T_s$ を出力する。それ以外の場合は、第1ユーザと第2ユーザの何れか一方が小さな声で話している状況である為、重複区間が発生していないものとして判定し、重複区間の発生時刻 $T_s$ をリセットする。なお、任意の閾値TH\_P\_DIFFは、例えば、3dBと規定すれば良い。なお、規定部4は、次式に基づいて重複区間を算出することが出来る。

40

(数12)

50

$$\begin{aligned} & \text{if } \{L(t) \geq TH\_L\} \wedge \{P1(t) - P2(t) < TH\_P\_DIFF\} & T_s = t - TH\_L \\ & \text{else} & T_s = -1 \end{aligned}$$

上述の(数12)において、 $T_s = t - TH\_L$ は重複区間の発生が有ることを意味し、 $T_s = -1$ は重複区間の発生が無いことを示す。なお、音声処理装置1は、実施例1ないし実施例4の処理を任意に組み合わせた音声処理を行うことが出来る。実施例4における音声処理装置1によれば、第2ユーザの第1ユーザに対する発話の印象をより高い精度で評価することが可能となる。

10

【0037】

(実施例5)

検出部3は、第1音声と第2音声に含まれる複数のフレームの第1信号品質と第2信号品質をそれぞれ検出し、規定部4は、第1信号品質と第2信号品質が第3閾値(例えば、第3閾値 = 10 dB)以上であり、かつ、第1発話区間と第2発話区間が重複する区間を重複区間として規定しても良い。なお、第1信号品質または第2信号品質は、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。例えば、検出部3は、第1発話区間に関する判定結果 $v_1(t)$ に応じて、次式に基づく雑音パワー $N1(t)$ を更新する。例えば、 $t$ 番目のフレームが非発話区間と判定された場合は、検出部3は、フレームパワーに忘却係数 $COEF1$ を乗じた値を加算することで、雑音パワーを更新する。なお、忘却係数 $COEF1$ は、例えば0.05と規定することが出来る。

20

(数13)

$$\begin{aligned} N1(t) &= N1(t-1) * COEF1 + P1(t) * (1 - COEF1) \\ N1(t) &= N1(t-1) \end{aligned}$$

上述の(数13)において、上段は $v_1(t) = 0$ の場合であり、下段は上記以外の場合を示す。また、検出部3は、第2発話区間に関する判定結果 $v_2(t)$ に応じて、雑音パワー $N1(t)$ と同様に雑音パワー $N2(t)$ を更新する。次に、規定部4は、重複区間の長さが上述の第1閾値 $TH\_L$ 以上であり、第1発話区間の第1信号品質の一例となる $SNR1(t)$ と、第2発話区間の第2信号品質の $SNR2(t)$ が第3閾値 $TH\_SNR$ (例えば10 dB)以上である場合に、重複区間の発生を検出し、重複区間の発生時刻 $T_s$ を次式に基づいて規定する。

30

(数14)

$$\begin{aligned} & \text{if } \{L(t) \geq TH\_L\} \wedge \{SNR1(t) \geq TH\_SNR\} \wedge \{SNR2(t) \geq TH\_SNR\} & T_s = t - TH\_L \\ & \text{else} & T_s = -1 \end{aligned}$$

なお、上述の(数14)において、 $T_s = t - TH\_L$ は重複区間の発生が有ることを意味し、 $T_s = -1$ は重複区間の発生が無いことを示す。

40

【0038】

また、規定部4は、第1信号品質と第2信号品質の差が第4閾値(例えば、第4閾値 = 3 dB)未満であり、かつ、第1発話区間と第2発話区間が重複する区間を重複区間として規定しても良い。規定部4は、次式に基づいて第1発話区間の第1信号品質の一例となる $SNR1(t)$ と第2発話区間の第2信号品質の一例となる $SNR2(t)$ の差と、第4閾値との比較結果に基づき、重複区間の発生を判定し、重複区間の発生時刻 $T_s$ を規定する。規定部4は、重複区間の長さが第1閾値 $TH\_L$ 以上であり、第1発話区間の $SNR1(t)$ と第2発話区間の $SNR2(t)$ との差が第4閾値 $TH\_P\_DIFF$ (例えば3 dB)未満である場合は、第1ユーザと第2ユーザともに同等程度の音声の大きさで

50

会話している状況である為、発話区間の重複が発生したと判定し、重複区間の発生時刻  $T_s$  を規定することが出来る。それ以外の場合は、第1ユーザと第2ユーザの何れか一方が小さな声で発話している状況であるため、重複区間が発生していないと判定し、重複区間の発生時刻  $T_s$  をリセットする。

(数15)

$$\begin{aligned} & \text{if } \{L(t) \geq TH\_L\} \wedge \{SNR1(t) - SNR2(t) < TH\_SNR\_DIFF\} & T_s = t - TH\_L \\ & \text{else} & T_s = -1 \end{aligned}$$

10

上述の(数15)において、 $T_s = t - TH\_L$ は重複区間の発生が有ることを意味し、 $T_s = -1$ は重複区間の発生が無いことを示す。なお、音声処理装置1は、実施例1ないし実施例5の処理を任意に組み合わせた音声処理を行うことが出来る。実施例5における音声処理装置1によれば、第2ユーザの第1ユーザに対する発話の印象をより高い精度で評価することが可能となる。

【0039】

(実施例6)

図7は、一つの実施形態による携帯端末装置30として機能するハードウェア構成図である。携帯端末装置30は、アンテナ31、無線部32、ベースバンド処理部33、端末インタフェース部34、マイク35、スピーカ36、制御部37、主記憶部38、補助記憶部39を有する。

20

【0040】

アンテナ31は、送信アンプで増幅された無線信号を送信し、また、基地局から無線信号を受信する。無線部32は、ベースバンド処理部33で拡散された送信信号をD/A変換し、直交変調により高周波信号に変換し、その信号を電力増幅器により増幅する。無線部32は、受信した無線信号を増幅し、その信号をA/D変換してベースバンド処理部33に伝送する。

【0041】

ベースバンド処理部33は、送信データの誤り訂正符号、データ変調、受信信号、受信環境の判定、各チャンネル信号の閾値判定、誤り訂正復号などのベースバンド処理などを行う。

30

【0042】

制御部37は、例えば、CPU、MPU(Micro Processing Unit)、DSP(Digital Signal Processor)、ASIC(Application Specific Integrated Circuit)、またはPLD(Programmable Logic Device)である。制御部37は、制御信号の送受信などの無線制御を行う。また、制御部37は、補助記憶部39などに記憶されている音声処理プログラムを実行し、例えば、実施例1または実施例2における音声処理を行う。換言すると、制御部37は、例えば、図1に記載の取得部2、検出部3、規定部4、算出部5、評価部6の機能ブロックの処理を実行することが出来る。

40

【0043】

主記憶部38は、ROMやRAMなどであり、制御部37が実行する基本ソフトウェアであるOSやアプリケーションソフトウェアなどのプログラムやデータを記憶又は一時保存する記憶装置である。

【0044】

補助記憶部39は、HDDやSSDなどであり、アプリケーションソフトウェアなどに関連するデータを記憶する記憶装置である。

【0045】

端末インタフェース部34は、データ用アダプタ処理、ハンドセットおよび外部データ端末とのインタフェース処理を行う。

50

## 【0046】

マイク35は、発話者（例えば、第1ユーザ）の音声を入力し、マイク信号として制御部37に出力する。スピーカ36は、出力音声または制御信号として制御部37から出力された信号を出力する。

## 【0047】

（実施例7）

図8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8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。

## 【0048】

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(Micro 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わせであってもよい。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、規定部4、算出部5、評価部6等の機能ブロックの処理を実行することが出来る。

## 【0049】

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。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。

## 【0050】

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。

## 【0051】

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。

## 【0052】

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。

## 【0053】

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(Digi

10

20

30

40

50

tal 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。

#### 【0054】

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やメモリリーダーライタ115を接続することが出来る。メモリ装置114は、機器接続  
10  
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。メモリリーダーライタ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35やスピーカ36を接続することが出来る。

#### 【0055】

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。

#### 【0056】

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  
20  
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、規定部4、算出部5、評価部6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  
30  
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。

#### 【0057】

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。  
40

#### 【0058】

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ユーザの第1音声に含まれる第1発話区間と、第2ユーザの第2音声に含まれる第2発話区間を検出する検出部と、

前記第1発話区間と前記第2発話区間が重複する重複区間を規定する規定部と、

前記重複区間の始点から前記第1発話区間の終点までの第1発話継続区間を算出する算出部と、

少なくとも前記第1発話継続区間の長さに関する情報に基づいて、前記第1音声に対する印象を評価する評価部

10

20

30

40

50

を備えることを特徴とする音声処理装置。

(付記 2)

前記規定部は、前記第 1 発話区間と前記第 2 発話区間が第 1 閾値以上の長さで重複する場合に、前記重複区間を規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)

前記規定部は、前記第 1 発話区間または前記第 2 発話区間に含まれる母音数を算出し、前記第 1 発話区間と前記第 2 発話区間が第 2 閾値以上の前記母音数で重複する場合に、前記重複区間を規定することを特徴とする付記 1 または付記 2 記載の音声処理装置。

(付記 4)

前記検出部は、前記第 1 音声と前記第 2 音声に含まれる複数のフレームの第 1 信号品質と第 2 信号品質をそれぞれ検出し、

前記規定部は、前記第 1 信号品質と前記第 2 信号品質が第 3 閾値以上であり、かつ、前記第 1 発話区間と前記第 2 発話区間が重複する区間を前記重複区間として規定する付記 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 音声と前記第 2 音声に含まれる複数のフレームの第 1 信号品質と第 2 信号品質をそれぞれ検出し、

前記規定部は、前記第 1 信号品質と前記第 2 信号品質の差が第 4 閾値未満であり、かつ、前記第 1 発話区間と前記第 2 発話区間が重複する区間を前記重複区間として規定する付記 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 6)

前記第 1 信号品質または前記第 2 信号品質は、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 3 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)

前記規定部は、前記第 1 発話区間と、前記第 2 発話区間のいずれか早い区間の開始時刻から第 5 閾値の時間経過後に前記重複区間を規定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)

前記規定部は、前記第 1 発話区間と、前記第 2 発話区間のいずれか遅い区間の終了時刻から前記第 5 閾値の時間経過前に前記重複区間を規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)

前記第 1 音声または前記第 2 音声は、送話音声または受話音声であることを特徴とする付記 1 ないし付記 8 の何れか一つに記載の音声処理装置。

(付記 10)

前記評価部は、前記第 1 発話継続区間の長さが第 6 閾値以上の場合に、前記印象を悪い印象と評価することを特徴とする付記 1 ないし付記 9 の何れか一つに記載の音声処理装置。

(付記 11)

第 1 ユーザの前記第 1 音声に含まれる第 1 発話区間と、第 2 ユーザの第 2 音声に含まれる第 2 発話区間を検出し、

前記第 1 発話区間と前記第 2 発話区間が重複する重複区間を規定し、

前記重複区間の始点から前記第 1 発話区間の終点までの第 1 発話継続区間を算出し、少なくとも前記第 1 発話継続区間の長さに関する情報に基づいて、前記第 1 音声に対する印象を評価する

ことを含むことを特徴とする音声処理方法。

(付記 12)

前記規定することは、前記第 1 発話区間と前記第 2 発話区間が第 1 閾値以上の長さで重複する場合に、前記重複区間を規定することを特徴とする付記 11 記載の音声処理方法。

10

20

30

40

50

## (付記 13)

前記規定することは、前記第 1 発話区間または前記第 2 発話区間に含まれる母音数を算出し、

前記第 1 発話区間と前記第 2 発話区間が第 2 閾値以上の前記母音数で重複する場合に、前記重複区間を規定することを特徴とする付記 11 または付記 12 記載の音声処理方法。

## (付記 14)

前記検出することは、前記第 1 音声と前記第 2 音声に含まれる複数のフレームの第 1 信号品質と第 2 信号品質をそれぞれ検出し、

前記規定することは、前記第 1 信号品質と前記第 2 信号品質が第 3 閾値以上であり、かつ、前記第 1 発話区間と前記第 2 発話区間が重複する区間を前記重複区間として規定する付記 11 ないし付記 13 の何れか一つに記載の音声処理方法。

10

## (付記 15)

前記検出することは、前記第 1 音声と前記第 2 音声に含まれる複数のフレームの第 1 信号品質と第 2 信号品質をそれぞれ検出し、

前記規定することは、前記第 1 信号品質と前記第 2 信号品質の差が第 4 閾値未満であり、かつ、前記第 1 発話区間と前記第 2 発話区間が重複する区間を前記重複区間として規定する付記 11 ないし付記 13 の何れか一つに記載の音声処理方法。

## (付記 16)

前記第 1 信号品質または前記第 2 信号品質は、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 ないし付記 15 の何れか一つに記載の音声処理方法。

20

## (付記 17)

前記規定することは、前記第 1 発話区間と、前記第 2 発話区間のいずれか早い区間の開始時刻から第 5 閾値の時間経過後に前記重複区間を規定することを特徴とする付記 11 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)

前記規定することは、前記第 1 発話区間と、前記第 2 発話区間のいずれか遅い区間の終了時刻から前記第 5 閾値の時間経過前に前記重複区間を規定することを特徴とする付記 11 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)

前記第 1 音声または前記第 2 音声は、送話音声または受話音声であることを特徴とする付記 11 ないし付記 18 の何れか一つに記載の音声処理方法。

30

## (付記 20)

前記評価することは、前記第 1 発話継続区間の長さが第 6 閾値以上の場合に、前記印象を悪い印象と評価することを特徴とする付記 11 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)

コンピュータに、

第 1 ユーザの第 1 音声に含まれる第 1 発話区間と、第 2 ユーザの第 2 音声に含まれる第 2 発話区間を検出し、

40

前記第 1 発話区間と前記第 2 発話区間が重複する重複区間を規定し、

前記重複区間の始点から前記第 1 発話区間の終点までの第 1 発話継続区間を算出し、

少なくとも前記第 1 発話継続区間の長さに関する情報に基づいて、前記第 1 音声に対する印象を評価する

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】

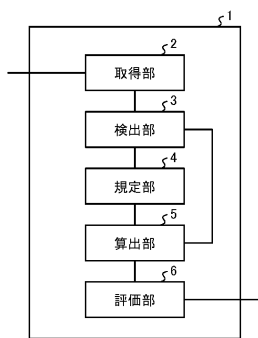
## 【0059】

- 1 音声処理装置
- 2 取得部
- 3 検出部

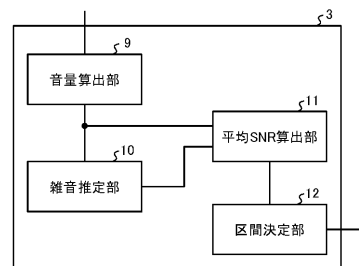
50

- 4 規定部
- 5 算出部
- 6 評価部

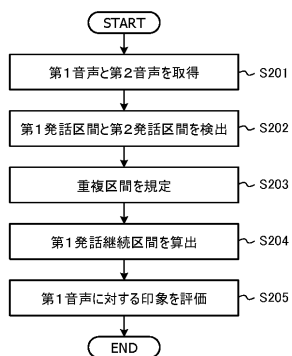
【図1】



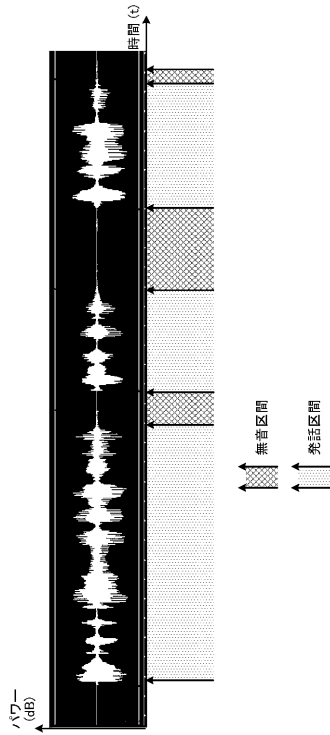
【図3】



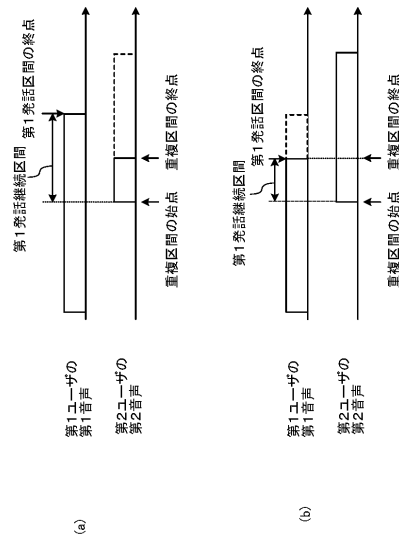
【図2】



【図4】



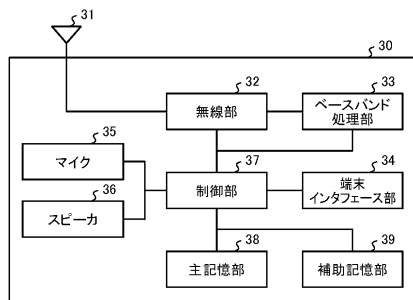
【図5】



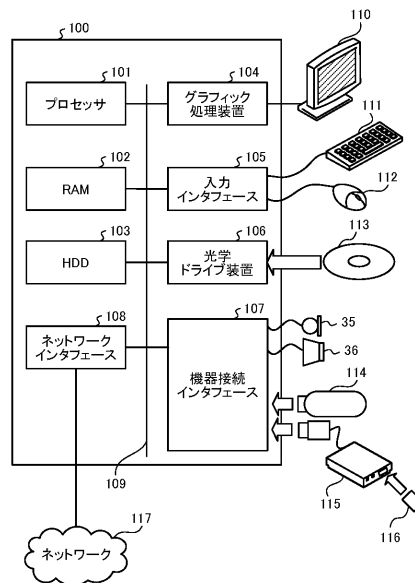
【図6】

第1発話継続区間	評価結果
0秒以上3秒未満	普通
3秒以上4秒未満	やや悪い
4秒以上	非常に悪い

【図7】



【図8】



---

フロントページの続き

(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

審査官 安田 勇太

(56)参考文献 特開2010-175684(JP,A)

特開2011-238028(JP,A)

国際公開第2011/105003(WO,A1)

米国特許出願公開第2012/0296642(US,A1)

特開平08-036396(JP,A)

(58)調査した分野(Int.Cl., DB名)

G10L 13/00 - 25/93

H04M 1/00