

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第6464703号
(P6464703)

(45) 発行日 平成31年2月6日(2019.2.6)

(24) 登録日 平成31年1月18日(2019.1.18)

(51) Int. Cl.		F I	
G 1 0 L 25/63	(2013.01)	G 1 0 L	25/63
G 1 0 L 25/60	(2013.01)	G 1 0 L	25/60
G 1 0 L 25/90	(2013.01)	G 1 0 L	25/90
G 1 0 L 15/10	(2006.01)	G 1 0 L	15/10 5 0 0 N

請求項の数 5 (全 18 頁)

(21) 出願番号	特願2014-243327 (P2014-243327)	(73) 特許権者	00004075
(22) 出願日	平成26年12月1日(2014.12.1)		ヤマハ株式会社
(65) 公開番号	特開2016-105142 (P2016-105142A)		静岡県浜松市中区中沢町10番1号
(43) 公開日	平成28年6月9日(2016.6.9)	(74) 代理人	100125689
審査請求日	平成29年10月23日(2017.10.23)		弁理士 大林 章
		(74) 代理人	100121108
			弁理士 高橋 太郎
		(72) 発明者	嘉山 啓
			静岡県浜松市中区中沢町10番1号 ヤマハ株式会社内
		審査官	大野 弘

最終頁に続く

(54) 【発明の名称】 会話評価装置およびプログラム

(57) 【特許請求の範囲】

【請求項1】

問いのうち特定区間の音高を前記問いの音声特徴として取得する第1音高取得部と、問いに対する回答の音高を前記回答の音声特徴として取得する第2音高取得部とを備える解析部と、

少なくとも前記第1音高取得部で取得された前記問いの音高及び前記第2音高取得部で取得された前記回答の音高に基づいて、前記問いに対する前記回答を評価する評価部と、を具備することを特徴とする会話評価装置。

【請求項2】

前記評価部は、

前記第1音高取得部で取得された前記問いの音高と前記第2音高取得部で取得された前記回答の音高との差分値が所定の範囲内に入るか否かを判定し、

前記所定の範囲内に入らない場合は、前記所定の範囲内に入るように前記回答の音高の音高シフト量をオクターブ単位で決定し、

前記回答の音高を前記音高シフト量だけシフトしたシフト後の回答の音高を、前記回答の音高として処理する、

ことを特徴とする請求項1に記載の会話評価装置。

【請求項3】

前記評価部は、前記問いの音高から前記回答の音高を減算した音高減算値が所定の基準値からどれだけ離れるかによって、前記問いに対する前記回答を評価する、ことを特徴と

する請求項 2 に記載の会話評価装置。

【請求項 4】

前記問いが終了してから前記回答が開始するまでの時間である会話間隔を検出する会話間隔検出部を備え、

前記評価部は、

前記第 1 音高取得部で取得された問いの音高及び前記第 2 音高取得部で取得された回答の音高、並びに前記会話間隔に基づいて、前記問いに対する前記回答を評価する、

ことを特徴とする請求項 1 乃至 3 のうちいずれかに記載の会話評価装置。

【請求項 5】

コンピュータを、

問いのうち特定区間の音高を前記問いの音声特徴として取得する第 1 音高取得部と、問いに対する回答の音高を前記回答の音声特徴として取得する第 2 音高取得部とを備える解析部と、

少なくとも前記第 1 音高取得部で取得された前記問いの音高及び前記第 2 音高取得部で取得された前記回答の音高に基づいて、前記問いに対する前記回答を評価する評価部、

として機能させることを特徴とするプログラム。

10

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、会話評価装置およびプログラムに関する。

20

【背景技術】

【0002】

従来、話し手が発言した音声自体を分析することで、話し手の心理状態などを分析するものが提案されている。例えば特許文献 1 では、話し手の音声シーケンスを取得し、その音声シーケンス中にある基音（1つの基本トーン：fundamental tone）の間隔や音程を検出することで、話し手の心理状態や健康状態などを診断する技術が提案されている。

【先行技術文献】

【特許文献】

30

【0003】

【特許文献 1】特許第 4 4 9 5 9 0 7 号公報

【発明の概要】

【発明が解決しようとする課題】

【0004】

ところで、人同士の会話では、相手から問いが発言されたとき、それに対して相槌を含め何らかの回答を発言する。このとき、どのように回答するかによって、相手に与える印象が異なる。

【0005】

これに対して、上述した特許文献 1 の技術は、1 人の話し手の音声シーケンスの中での基音間隔や音程によって、話し手の心理状態などを分析するものである。したがって、2 人の会話中の問いと回答の音声特徴の比較で、その問いに対する回答を評価するものではない。このため、特許文献 1 の技術では、会話中における回答が、問いに対する回答として良好かどうか評価することはできない。

40

【0006】

本発明は、このような事情に鑑みてなされたものであり、その目的の一つは、回答の音声特徴を問いの音声特徴との比較で評価することで、その問いに対する回答として相手に与える印象を客観的に確認できる会話評価装置およびプログラムを提供することにある。

【課題を解決するための手段】

【0007】

50

このような会話中の問いに対する回答を評価するに当たって、まず人同士でどのような会話（対話）がなされるかについて、言語的情報以外の情報、とりわけ対話の特徴付ける音高（周波数）に着目して考察する。

【0008】

人同士の対話として、一方の人（aとする）による問い（問い掛け）に対し、他方の人（bとする）が回答（返答）する場合について検討する。この場合において、aが問いを発したとき、aだけでなく、当該問いに対して回答しようとするbも、当該問いのうちの、特定区間における音高を強い印象で残していることが多い。bは、同意や、賛同、肯定などの意で回答するときには、印象に残っている問いの音高に対し、当該回答を特徴付ける部分の音高が、特定の関係、具体的には協和音程の関係となるように発声する。当該回答を聞いたaは、自己の問いについて印象に残っている音高と当該問いに対する回答を特徴付ける部分の音高とが上記関係にあるので、bの回答に対して心地良く、安心するような好印象を抱くことになる、と考えられる。

10

【0009】

このように人同士の対話では、問いの音高と回答の音高とは無関係ではなく、上記のような関係がある、と考察できる。このような考察を踏まえて、問いに対する回答を評価する会話評価システムを検討したときに、上記目的を達成するために、次のような構成とした。

【0010】

すなわち、上記目的を達成するために、本発明の一態様に係る会話評価装置は、問いのうち特定区間の音高を前記問いの音声特徴として取得する第1音高取得部と、問いに対する回答の音高を前記回答の音声特徴として取得する第2音高取得部とを備える解析部と、少なくとも前記第1音高取得部で取得された前記問いの音高及び前記第2音高取得部で取得された前記回答の音高に基づいて、前記問いに対する前記回答を評価する評価部と、を具備することを特徴とする。

20

【0011】

この一態様によれば、問いに対する回答の音声特徴としての音高を問いの音声特徴としての音高との比較で評価することができる。これにより、その問いに対する回答として相手に与える印象を客観的に確認することができる。

【0012】

上述したように問いの音高と回答の音高とは、相手に与える印象に密接な関係があるので、回答の音高を問いの音高との比較で評価することで、問いに対する回答について信頼性の高い評価をすることができる。

30

【0013】

上記態様において、前記第1音高取得部で取得された前記問いの音高と前記第2音高取得部で取得された前記回答の音高との差分値が所定の範囲内に入るか否かを判定し、

前記所定の範囲内に入らない場合は、前記所定の範囲内に入るように前記回答の音高の音高シフト量をオクターブ単位で決定し、前記回答の音高を前記音高シフト量だけシフトしたシフト後の回答の音高を、前記回答の音高として処理するようにしてもよい。これによれば、男性と女性の会話や大人と子どもの会話のように、問いと回答の発話の音程が大きく異なる場合においても、問いに対する回答を適正に評価できる。この場合、前記問いの音高から前記回答の音高を減算した音高減算値が所定の基準値からどれだけ離れるかによって、前記問いに対する前記回答を評価することができる。

40

【0014】

上記態様において、前記問いが終了してから前記回答が開始するまでの時間である会話間隔を検出する会話間隔検出部を備え、前記評価部は、前記第1音高取得部で取得された問いの音高及び前記第2音高取得部で取得された回答の音高、並びに前記会話間隔に基づいて、前記問いに対する前記回答を評価するようにしてもよい。問いに対する回答の音声特徴として、上述した音高の他にも、問いの終了から回答の開始までの時間（会話間隔）は相手に与える印象に密接な関係がある。このため、問いと回答の音高のみならず、問い

50

と回答の会話間隔についても評価することで、問いに対する回答についてより信頼性の高い評価をすることができる。

【0015】

本発明の態様について、会話評価装置のみならず、コンピュータを当該会話評価装置として機能させるプログラムとして概念することも可能である。

【図面の簡単な説明】

【0016】

【図1】本発明の第1実施形態に係る会話評価装置の構成を示すブロック図である。

【図2】図1に示す会話評価装置の動作の一例を示すメインルーチンのフローチャートである。

10

【図3】図2に示す会話の評価を行う際のサブルーチンを示すフローチャートである。

【図4】本実施形態における問いと回答との音高例を示す図である。

【図5】本実施形態における問いと回答との音高例を示す図であって、問いと回答との音高差分値が1オクターブ以上ある場合の例を示す。

【図6】本実施形態における音高評価点の算出基準の具体例を説明するための図である。

【図7】本実施形態における会話間隔評価点の算出基準の具体例を説明するための図である。

【図8】本発明の第2実施形態に係る会話評価装置の構成を示すブロック図である。

【図9】図8に示す会話評価装置の動作の一例を示すメインルーチンのフローチャートである。

20

【図10】本発明の第3実施形態に係る会話評価装置の構成を示すブロック図である。

【図11】図10に示す会話評価装置の動作の一例を示すメインルーチンのフローチャートである。

【発明を実施するための形態】

【0017】

以下、本発明の実施形態について図面を参照して説明する。

【0018】

<第1実施形態>

図1は、本発明の第1実施形態に係る会話評価装置10の構成を示す図である。ここでの会話評価装置10は、2人の会話音声を1つの音声入力部のマイクロフォンで入力し、会話中の問いに対する回答を評価して表示する会話トレーニング装置に適用した場合を例に挙げる。またここでの問いに対する回答には、問いの質問に答える回答のみならず、例えば「はい」、「いいえ」、「そう」、「うん」、「ふーん」、「なるほど」のような質問に対する返事や相槌（間投詞）も含まれる。

30

【0019】

図1に示すように、会話評価装置10は、CPU（Central Processing Unit）、メモリやハードディスク装置などの記憶部、1つの音声入力部102、表示部112などを有し、当該CPUが、予めインストールされたアプリケーションプログラムを実行することによって、複数の機能ブロックが次のように構築される。詳細には、会話評価装置10では、音声取得部104、解析部106、判別部108、言語データベース122、会話間隔検出部109および評価部110が構築される。

40

【0020】

なお、特に図示しないが、このほかにも会話評価装置10は、操作入力部などを備え、利用者が装置に対して各種の操作を入力し、各種の設定などができるようになっている。また、会話評価装置10は、会話トレーニング装置に限られず、スマートフォンや携帯電話機のような端末装置やタブレット型のパーソナルコンピュータなどであっても良い。また、3人以上の会話音声を1つの音声入力部102のマイクロフォンで入力する場合に適用してもよい。この場合、例えば1人が問いを発話したときに、その問いに対する回答は、他の2人のうちの誰が回答してもよい。

【0021】

50

音声入力部 102 は、詳細については省略するが、音声を電気信号に変換するマイクロフォンと、変換された音声信号をデジタル信号に変換する A/D 変換器とで構成される。

【0022】

音声取得部 104 は、リアルタイムでデジタル信号に変換された音声信号を取得してその音声信号を一時的にメモリに記憶する。

【0023】

解析部 106 は、デジタル信号に変換された音声信号の解析処理を行って発話（問いや回答）の音声特徴（音高や音量など）を抽出する。解析部 106 は、問いのうち特定区間の音高（ピッチ）を問いの音声特徴として取得する第 1 音高取得部 106A と、回答の音声に基づく音高を回答の音声特徴として取得する第 2 音高取得部 106B とを備える。

10

【0024】

第 1 音高取得部 106A は、問いの音声信号において発話開始から発話終了までの発話区間のうち、有声区間における特定区間の音高を検出し、当該音高を示すデータを評価部 110 に供給する。ここでの特定区間は、発話が終了する直前の所定時間の末尾区間（例えば 180 msec）であり、当該末尾区間における最高値を音高として検出する。

【0025】

本実施形態のようにリアルタイムで音声を入力する場合、発話開始は例えば音声信号の音量が閾値以上になったことで判断することができ、発話終了は例えば音声信号の音量が一定期間閾値未満となったことで判断することができる。なお、チャタリングを防止するため、複数の閾値を用い、ヒステリシス特性を付与してもよい。また、有声区間とは、発話区間のうち、音声信号の音高（ピッチ）が検出可能な区間をいう。音高が検出可能な区間とは、音声信号に周期的な部分があって、その部分が検出可能であることを意味する。

20

【0026】

なお、問いの有声区間の末尾区間が無声音（端的に言えば、発声の際に声帯の振動を伴わない音）である場合、直前の有声音部分から、当該無声音部分の音高を推定しても良い。問いの特定区間については、有声区間の末尾区間に限られるものではなく、例えば語頭区間であっても良い。また、問いのうちのどの部分の音高を特定するかについて、利用者が任意に設定できる構成としても良い。また、有声区間の検出のために音量および音高の 2 つを用いるのではなく、いずれか一方を用いて検出しても良いし、どれを用いて有声区間の検出をするのかを利用者が選択しても良い。

30

【0027】

第 2 音高取得部 106B は、回答の音声信号から音高（例えば発話区間の平均音高）を検出し、当該音高を示すデータを評価部 110 に供給する。

【0028】

解析部 106 は、音声取得部 104 でメモリに記憶された音声信号を用いて、特定区間の検出やその特定区間の音高を検出してもよく、リアルタイムの音声信号を用いて音高を検出してもよい。リアルタイムで問いの音高を検出する場合には、例えば入力した音声信号の音高を、直前の音声信号の音高と比較して高い方の音高を記憶して更新する。これを問いの発話終了まで続けることで、最終的に更新された音高を問いの音高として特定する。これにより、発話終了まで最大の音高を問いの音高として特定できる。また、回答の音高を検出する場合は、音節によって特定してもよい。例えば相槌の回答の場合は第 2 音節あたりの音高が全体の平均に近くなることが多いので、第 2 音節開始時の音高を回答の音高として特定するようにしてもよい。

40

【0029】

判別部 108 は、デジタル信号に変換された発話の音声信号を解析し、文字列に変換する音声認識を行うことで、発話の言葉の意味を特定する。これにより、その発話が問いか回答かを判別し、判別結果を示すデータを解析部 106 に供給する。判別部 108 は、発話の意味を特定する際に、その発話の音声信号がどの音素に近いのかを、言語データベース 122 に予め作成された音素モデルを参照することにより判定して、音声信号で規定される言葉の意味を特定する。このような音素モデルには、例えば隠れマルコフモデルを用

50

いることができる。

【0030】

なお、判別部108による発話の判別は、上記の方法に限られるのではなく、音声特徴の変化によって行うようにしてもよい。例えば語尾区間の音高が上昇した発話があればそれは問いと判別でき、その次の発話の音声は2音節であれば相槌の回答と判別できる。また、通常は発話が問いであれば、次の発話は回答である。このため、判別部108では、少なくとも発話が問いか否かを判別できればよい。

【0031】

ところで、人同士の対話において問いに対して回答する場合、音高以外にも考慮される要素として、問いの終了から回答の開始までの時間（会話間隔）がある。例えば、二択で回答を迫るような問いに対して「いいえ」と回答する場合、慎重を期するために、一呼吸遅れるように間を取る点も、経験上よく見られる行為である。

10

【0032】

人同士の対話において、二択ではなく、例えばWho（誰が）、What（何を）、When（いつ）、Where（どこで）、Why（なぜ）、How（どのようにして）のような5W1Hの問いに対しては、ゆっくりと時間をかけて具体的内容を回答する場合がある。

【0033】

いずれの場合でも、問いの終了から回答の開始までの時間が空くと、問いを発話した相手に一種の不安感を与えてしまうとともに、以降の会話が弾まない。また、逆に回答までの間が詰まり過ぎると、意識的に被されているかのような感覚、または、人の話をまともに聞いていないのではないかという感覚になり、不快感を与えてしまう。

20

【0034】

そこで、本実施形態では、問いに対する回答の評価を行う際に、音高だけではなく、問いの終了から回答の開始までの会話間隔を測定して、これを評価できるようにしている。詳細には、会話間隔検出部109において、問いの終了から回答の開始までの時間（会話間隔）を検出する。会話間隔は、会話評価装置10に内蔵されるタイマまたはリアルタイムクロックで計時する。タイマで計時する場合には、問いの終了により計時を開始し、回答の開始により計時を終了することで、その間の時間を会話間隔として検出する。リアルタイムクロックで計時する場合には、問いの終了時と回答の開始時の時刻を取得しておき、その間の時間を会話間隔として検出する。検出された会話間隔の時間データは、評価部110に供給され、上述した問いと回答の音高データとともに評価の対象とされる。

30

【0035】

評価部110は、解析部106からの問いと回答の音高データと、会話間隔検出部109からの時間データにより、問いに対する回答の評価を行って評価点を算出する。詳細には、音高データの評価は、問いの音高から回答の音高を減算した音高減算値を求め、この音高減算値が所定の基準値からどれだけ離れているかという観点から音高評価点を算出する。会話間隔の時間データの評価は、会話間隔の時間が所定の基準値からどれだけ離れているかという観点から会話間隔評価点を算出する。評価部110は、これら音高評価点と会話間隔評価点の合計を最終的な回答の評価点として算出し、表示部112に表示する。これにより、回答者は、問いに対する回答の評価を確認することができる。なお、評価部110による評価の詳細は後述する。

40

【0036】

次に、会話評価装置10の動作について説明する。図2は、会話評価装置10における処理動作を示すフローチャートである。はじめに、利用者が所定の操作をしたとき、例えば当該対話のための処理に対応したアイコンなどをメインメニュー画面（図示省略）において選択したとき、CPUが当該処理に対応したアプリケーションプログラムを起動する。このアプリケーションプログラムを実行することによって、CPUは、図1で示した機能ブロックを構築する。

【0037】

50

ここでは、1つの音声入力部102のマイクロフォンで2人の自然の会話の音声を入力し、リアルタイムで音声特徴を取得しながら、問いに対する回答の評価を行う場合を例にとって説明する。このように自然の会話を1つの音声入力部102で入力する場合には、発話が問いか回答か不明なため、発話が問いか否かの判別が必要となる。なお、ここでは説明の便宜のため、発話が問いであると判別されれば、その後の発話は回答とし、その発話が回答であるか否かの判別は行わない。ただし、これに限られるものではなく、問いの後の発話が回答であるか否かについて判別するようにしてもよい。

【0038】

まず、ステップS a 1 1において、音声入力部102によって変換された音声信号が音声取得部104を介して解析部106に供給され、発話が開始されたか否かが判断される。例えば発話が開始されたか否かは、音声信号の音量が閾値以上になったか否かで判断される。なお、音声取得部104は音声信号をメモリに記憶する。

10

【0039】

発話が開始されたと判断されると、ステップS a 1 2において、解析部106の第1音高取得部106Aにより、音声取得部104からの音声信号に対して発話の音高を音声特徴として取得する解析処理が行われる。ステップS a 1 1において発話が開始されたと判断されなければ、発話が開始されたと判断されるまでステップS a 1 1が繰り返される。

【0040】

ステップS a 1 3において、解析部106によって発話中か否かが判断される。発話中か否かは、閾値以上の音量の音声信号が続いているか否かで判断される。ステップS a 1 3において発話中であると判断されると、ステップS a 1 2に戻り、音高を取得するための解析処理が継続される。ステップS a 1 3において発話中でないと判断されると、ステップS a 1 4において判別部108により発話は問いか否かが判断される。ステップS a 1 4において発話は問いでないと判断されると、ステップS a 1 1に戻り、次の発話の開始待ちとなる。

20

【0041】

これに対して、ステップS a 1 4において発話は問いであると判断されると、ステップS a 1 5において、発話(問い)が終了したか否かを判断する。問いが終了したか否かは、例えば音声信号の音量が所定の閾値未満となった状態が所定時間継続したか否かで判断される。

30

【0042】

ステップS a 1 5において発話(問い)が終了していないと判断されると、ステップS a 1 2に戻り、音高を取得するための解析処理が継続される。第1音高取得部106Aは、音声信号の解析処理によって、発話(問い)の音高(例えば問いの語尾区間の最高音高)を音声特徴として取得すると、その問いの音高データを評価部110に供給する。

【0043】

ステップS a 1 5において発話(問い)が終了したと判断されると、ステップS a 1 6において、会話間隔検出部109により会話間隔の計時が開始される。

【0044】

次に、ステップS a 1 7において、回答が開始されたか否かが判断される。このときには既に問いの終了後であるため、次の発話は回答になる。このため、回答が開始されたか否かは、問いの終了後の音声信号の音量が閾値以上になったか否かで判断される。

40

【0045】

ステップS a 1 7において回答が開始されたと判断されると、ステップS a 1 8において、会話間隔検出部109により会話間隔の計時が終了される。これにより、問いの終了から回答の開始までの会話間隔の時間を計時することができる。会話間隔検出部109は計時した会話間隔の時間データを評価部110に供給する。

【0046】

ステップS a 1 9において、解析部106の第2音高取得部106Bにより、音声取得部104からの音声信号に対して回答の音高を音声特徴として取得する解析処理が行われ

50

る。

【0047】

ステップS a 2 0において、回答が終了したか否かを判断する。回答が終了したか否かは、例えば音声信号の音量が所定の閾値未満となった状態が所定時間継続したか否かで判断される。

【0048】

ステップS a 2 0において回答が終了していないと判断されると、ステップS a 1 9に戻り、音高を取得するための解析処理が継続される。第2音高取得部106Bは、音声信号の解析処理によって、回答の音高（例えば回答の平均音高）を音声特徴として取得すると、その回答の音高データを評価部110に供給する。ステップS a 2 0において発話（回答）が終了したと判断されると、ステップS a 2 1において、評価部110によって会話の評価が実行される。

10

【0049】

図3は、図2におけるステップS a 2 1の会話評価の処理の詳細を示すフローチャートである。

【0050】

まず、ステップS b 1 1において、評価部110は、第1音高取得部106Aから取得した問いの音高データと第2音高取得部106Bから取得した回答の音高データとに基づいて問いの音高と回答の音高との差分値（問いの音高から回答の音高を減算した音高減算値の絶対値）を算出する。

20

【0051】

ステップS b 1 2において、評価部110は、算出された音高差分値が所定の範囲内か否かを判断する。この音高差分値が所定の範囲外であると判断されると、ステップS b 1 3において、評価部110は、回答の音高の調整を行う。具体的には、評価部110は、上記音高差分値が所定の範囲内（例えば1オクターブの範囲内）に入るように、回答の音高の音高シフト量をオクターブ単位で決定する。評価部110は、回答の音高を音高シフト量だけ調整して、ステップS b 1 1に戻り、問いの音高とシフト後の回答の音高とにより音高差分値を算出し直す。これによれば、地声が高い音声の人（例えば女性や子供）と地声が低い音声の人（例えば男性）との会話のように、地声で1オクターブ以上の音高差があるような場合においても、その地声などの音高差を修正して、問いに対する回答を適正に評価できるようにしたものである。なお、上述した男性と女性の会話のみならず、男性同士の会話でも、また女性同士の会話においても、地声で1オクターブ以上の音高差がある場合もあるので、このような場合にも、問いに対する回答を適正に評価できる。

30

【0052】

なお、上記音高差分値が所定の範囲内（例えば1オクターブの範囲内）に入るまで、ステップS b 1 3において回答の音高を1オクターブずつ調整するようにしてもよい。また、ここでは、問いの音高はそのまま回答の音高の方を調整する場合を例に挙げたが、これに限られるものではなく、回答の音高はそのまま問いの音高の方を調整するようにしてもよい。

【0053】

ステップS b 1 2において、評価部110は、上記音高差分値が所定の範囲であると判断されると、ステップS b 1 4において、評価部110は、問いの音高から回答の音高を減算した音高減算値に基づいて音高の評価点を算出する。このとき、ステップS b 1 3において音高の調整を行った場合には、その音高の調整後の音高減算値を用いて音高の評価点を算出する。ここでの音高減算値は、問いの音高から回答の音高を減算したものであるから、回答の音高が問いの音高より低い場合はプラス値になり、回答の音高が問いの音高より高い場合はマイナス値になる。これは、回答の音高が問いの音高より低い場合を、問いの音高より高い場合よりも高評価にするためである。ステップS b 1 4における音高評価点は、上記音高減算値が所定の基準値からどれだけ離れているかという観点から算出される。例えば所定の基準値を700centとすれば、上記音高減算値が700cent

40

50

のときを満点(100点)とし、上記音高減算値が700centから離れるほど評価点の減算をすることで、問いに対する回答の音高評価点を算出する。これによれば、音高評価点が100点に近いほど、問いに対する回答が良好である。なお、上記音高減算値が所定の基準値に近づくほど評価点の加算をするようにしてもよい。

【0054】

次に、ステップS b 1 5において、評価部110は、会話間隔検出部109からの会話間隔の時間データに基づいて、会話間隔の評価点を算出する。このような会話間隔の評価は、問い終了から回答開始までの会話間隔の時間が所定の基準値からどれだけ離れているかという観点から算出される。例えば所定の基準値を180mscとすれば、会話間隔の時間が180mscのときを満点(100点)とし、会話間隔の時間が180mscから離れるほど評価点の減算をすることで、会話間隔評価点を算出する。これによれば、会話間隔評価点が100点に近いほど、問いに対する回答が良好である。なお、会話間隔の時間が所定の基準値に近づくほど評価点の加算をするようにしてもよい。

10

【0055】

続いて、ステップS b 1 6において、評価部110は、問いに対する回答の音高評価点と会話間隔評価点から総合評価点を算出する。総合評価点は、単純に音高評価点と会話間隔評価点を加算して算出する。なお、総合評価点は、音高評価点と会話間隔評価点に所定の重み付けを付加してから加算して算出してもよい。

【0056】

次に、ステップS b 1 7において、評価部110は、問いに対する回答の評価結果を表示部112に表示させて、図2のステップS a 2 1に戻る。評価結果は、総合評価点のみを表示させる。これにより、問いに対する回答の評価を、評価点というスコア値で客観的に確認することができる。なお、総合評価点だけでなく、音高評価点と会話間隔評価点とを表示させるようにしてもよい。

20

【0057】

また、問いに対する回答の評価結果の表示は、評価点のみならず、表示部112に評価点に応じたイルミネーションやアニメーションを表示するようにしてもよい。また、問いに対する回答の評価結果は、表示部112の画面表示だけに限られるものではない。例えば会話評価装置10を携帯端末に適用した場合には、その携帯端末の振動機能や音発生機能を利用して、評価点に応じた振動パターンで会話評価装置10を振動させたり、評価点に応じた音を発生させたりするようにしてもよい。

30

【0058】

また、会話評価装置10をぬいぐるみなどの玩具やロボットに適用した場合には、問いに対する回答の評価結果を、ぬいぐるみやロボットの動作で表すようにしてもよい。例えば評価点が高い場合には、ぬいぐるみやロボットにばんざい動作をさせることができ、評価点が高い場合には、ぬいぐるみやロボットにがっかり動作をさせることもできる。これにより、問いに対する回答による会話トレーニングをより楽しく行うことができる。

【0059】

ここで、本実施形態における評価部110が行う音高の調整(ステップS b 1 2、S b 1 3)について図面を参照しながらより詳細に説明する。ここでは、問いと回答の音高差分値が、1オクターブ以内である場合(音高を調整しない場合)と、1オクターブ以内でない場合(音高を調整する場合)とを比較しながら説明する。

40

【0060】

図4と図5はそれぞれ、音声入力された問いと回答との関係を、音高を縦軸にとり、時間を横軸にとって例示した図である。図4は音高差分値が1オクターブ以内である場合であり、図5は音高差分値が1オクターブ以内でない場合である。

【0061】

図4および図5において、符号Qで示される実線は、問いの音高変化を簡易的に直線で示している。符号dQは、この問いQにおける特定区間の音高(語尾区間の最高音高)である。また、図4において、符号Aで示される実線は、問いQに対する回答の音高変化を

50

簡易的に直線で示しており、符号 d_A はこの回答 A の平均音高である。符号 D は、問い Q の音高 d_Q と回答 A の音高 d_A との差分値である。なお、図 4 の符号 t_Q は問い Q の終了時刻であり、符号 t_A は回答 A の開始時刻である。符号 T は、 t_Q と t_A との間の時間であり、問い Q の終了から回答 A の開始までの時間に相当する。

【 0 0 6 2 】

図 5 において、符号 A' で示される点線は、回答 A の音高を 1 オクターブだけシフトさせた音高調整後の回答の音高変化を直線で示したものである。符号 $d_{A'}$ はこの音高調整後の回答 A' の平均音高である。符号 D' は、問いの音高 d_Q と音高調整後の回答 A' の音高 $d_{A'}$ との差分値である。

【 0 0 6 3 】

図 4 においては、音高差分値 D が 1 オクターブ (1 2 0 0 c e n t) 以内である場合である。この場合には、音高の調整は不要であるため、図 3 のステップ $S b 1 1$ で音高差分値 D が算出された後は、ステップ $S b 1 3$ が実行されずに、ステップ $S b 1 4$ にて問い Q の音高 d_Q から回答 A の音高 d_A を減算した音高減算値によって音高評価点が算出される。ここでの音高減算値は、回答 A の音高 d_A が問い Q の音高 d_Q よりも低いのでその音高差はプラス値となるため、音高差分値 D と同値になる。

【 0 0 6 4 】

これに対して、図 5 においては、音高差分値 D が 1 オクターブ (1 2 0 0 c e n t) を超える場合である。この場合には、音高の調整が必要となる。図 5 では、回答 A の音高が問い Q の音高よりも低い方に大きくずれているので、例えば地声が高い人の問い Q に対して、1 オクターブ以上地声が低い人が回答 A をしたような場合である。このように同じ音量で同じ音声を発した場合でも、地声で 1 オクターブ以上の音高差がある場合には、そのまま問いと回答の音高差で評価しても、地声の差異の分だけ評価点が大きくずれてしまい、適切な評価ができない可能性がある。そこで、本実施形態においては、図 3 のステップ $S b 1 3$ で回答 A の音高 d_A を、高い方に 1 オクターブ R だけシフトさせて、回答 A' の音高 $d_{A'}$ に調整する。このように、問い Q の音高 d_Q と調整後の回答の音高 $d_{A'}$ との音高差分値 D' は、1 オクターブ (1 2 0 0 c e n t) 以内にす。これにより、発話機構の影響を少なくすることができるので、適切な音高評価点を算出することができる。なお、音高調整は、音高が高い方にオクターブ単位でシフトする場合に限られず、音高が低い方にオクターブ単位でシフトするようにしてもよい。

【 0 0 6 5 】

次に、本実施形態における評価部 1 1 0 が行う音高評価点の算出 (ステップ $S b 1 4$) について図面を参照しながらより詳細に説明する。図 6 は、音高評価点の算出基準の具体例を説明するための図であり、横軸には問いと回答との音高減算値 D をとり、縦軸には音高評価点をとっている。図 6 において、符号 D_0 は、音高減算値の基準値であり、例えば 7 0 0 c e n t である。図 6 に示す実線は、音高評価点の算出基準線であり、音高減算値 D が高い方にも低い方にも、音高基準値 D_0 から離れるほど評価点が低くなるような直線で示したものである。音高評価点の算出基準線は、基準値 D_0 から所定範囲 (下限値 D_L ~ 上限値 D_H) 外は、音高評価点が 0 になるように設定されている。このため、例えば音高減算値が基準値 D_0 である場合を 1 0 0 点とすれば、所定範囲 (下限値 D_L ~ 上限値 D_H) 内において基準値 D_0 から離れるほど点数が低くなり、所定範囲 (下限値 D_L ~ 上限値 D_H) 外では 0 になる。なお、図 6 の音高評価点の算出基準線は、基準値 D_0 を通る縦軸に平行な直線に対して線対称となる場合を例に挙げているが、必ずしも線対称でなくてもよい。例えば基準値 D_0 の前後で直線の傾きを変えるようにしてもよい。また、音高評価点の算出基準線は、直線に限られるものではなく、曲線であってもよい。また音高評価点の算出基準線は、線形に限られず、非線形であってもよい。

【 0 0 6 6 】

図 6 に示す音高評価点の算出基準線によって音高評価点を算出する場合には、算出された問い Q の音高から回答 A の音高を減算した音高減算値を D_x とすれば、算出基準線で D_x に対応する $S d_x$ が音高評価点の加算点または減算点となる。例えば初期の音高評価点

10

20

30

40

50

を0点とすれば、その0点に加算点(減算点)を加算(減算)することによって、音高評価点を算出する。

【0067】

音高減算値の基準値D0は、問いに対する最適な回答の音高になるように設定することが好ましい。ここでは、基準値D0を700centに設定した場合を例に挙げている。これは、問いの音高に対して回答の音高が略5度下の関係、すなわち協和音程の関係になる音高減算値である。このように、基準値D0は、問いと回答の音高減算値が協和音程の関係になる音高減算値であることが好ましい。これは人同士の会話において、問いに対して完全肯定をする場合には、問いと回答の音高減算値が協和音程の関係に近いほど、心地良く、安心するような好印象を抱く適切な回答になるからである。これにより、問いの音高から回答の音高を減算した音高減算値が基準値に近いほど、問いに対して良好な回答であると評価できる。なお、問いの音高に対する回答の音高の関係は、上述した略5度下の協和音程の関係に限られるものではなく、略5度下以外の協和音程の関係としてもよい。例えば、完全8度、完全5度、完全4度、長・短3度、長・短6度であっても良い。さらに、協和音程の関係でなくても、経験的に良い印象を与える音程の関係の存在が認められる場合もあるので、当該音程の関係にしても良い。

10

【0068】

次に、本実施形態における評価部110が行う会話間隔評価点の算出(ステップSb15)について図面を参照しながらより詳細に説明する。図7は、会話間隔評価点の算出基準の具体例を説明するための図であり、横軸には会話間隔の時間Tをとり、縦軸には会話間隔評価点をとっている。図7において、符号T0は、会話間隔評価の基準値であり、例えば180msである。図7に示す実線は、会話間隔評価点の算出基準線であり、会話間隔の時間Tが長くなる方にも短くなる方にも、会話間隔基準値T0から離れるほど評価点が低くなるような直線で示したものである。会話間隔評価点の算出基準線は、基準値T0から所定範囲(下限値TL~上限値TH)外になると、会話間隔評価点が0になるように設定されている。このため、例えば会話間隔の時間が基準値T0である場合を100点とすれば、所定範囲(下限値TL~上限値TH)内において基準値T0から離れるほど点数が低くなり、所定範囲(下限値TL~上限値TH)外では0になる。なお、図7の会話間隔評価点の算出基準線は、基準値T0を通る縦軸に平行な直線に対して線対称となる場合を例に挙げているが、必ずしも線対称でなくてもよい。例えば基準値T0の前後で直線の傾きを変えるようにしてもよい。また、会話間隔評価点の算出基準線は、直線に限られるものではなく、曲線であってもよい。また会話間隔評価点の算出基準線は、線形に限られず、非線形であってもよい。

20

30

【0069】

図7に示す会話間隔評価点の算出基準線によって会話間隔評価点を算出する場合には、算出された問いQと回答Aの会話間隔時間をTxとすれば、算出基準線でTxに対応するStxが会話間隔評価点の加算点または減算点となる。例えば初期の会話間隔評価点を0点とすれば、その0点に加算点(減算点)を加算(減算)することによって、会話間隔評価点を算出する。

【0070】

会話間隔の基準値T0は、問い終了から回答開始までの最適な時間を設定することが好ましい。ここでは、基準値T0を180msに設定した場合を例に挙げている。これは問いに対する回答が相手に心地良く、安心するような好印象を抱かせる会話間隔の時間である。これによれば、問い終了から回答開始までの会話間隔の時間が、基準値に近いほど、問いに対して良好な回答であると評価できる。

40

【0071】

なお、音高減算値の基準値D0、会話間隔の時間の基準値T0は、必ずしも完全肯定の回答を評価する場合の基準値に限られるものではない。怒りの回答、気のない回答のような感情を伴った回答など回答の種類に応じて会話間隔の基準値T0を変更するようにしてもよい。これにより、問いに対する回答の種類に応じて、適切な回答の評価が可能となる

50

。例えば怒りの回答を評価する場合には、会話間隔の時間の基準値 T_0 を完全肯定の場合 (180ms) よりも短くする。これにより、問いに対する回答の怒りの度合いを評価することができる。また気のない回答を評価する場合には、会話間隔の時間の基準値 T_0 を完全肯定の場合 (180ms) よりも長くする。これにより、問いに対する回答の気のない度合いを評価することができる。

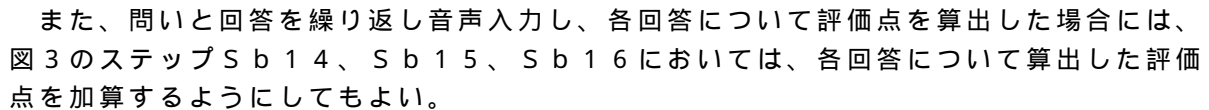
【0072】

また、音高減算値の基準値 D_0 、会話間隔の時間の基準値 T_0 は、上記のような回答の種類に応じて複数設けるようにしてもよい。例えば完全肯定の回答の場合の基準値、怒りの回答の場合の基準値、気のない回答の場合の基準値を別々に設けるようにしてもよい。

【0073】

また、問いと回答の音声特徴として、音高の他に音量についても評価するようにしてもよい。詳細には、例えば問いと回答の音量を音声特徴として取得し、問いの音量と回答の音量の差分値を求め、この差分値が所定の基準値からどれだけ離れているかという観点から音量評価点を算出する。音量評価点は、音高評価点と会話間隔評価点に加算して総合評価点を算出する。音量差分値の基準値についても、上記回答の種類に応じて変更したり、複数の基準値を設けたりしてもよい。例えば気のない回答の場合は、完全肯定の回答の場合よりも基準値を低くする。これにより、問いに対する回答の気のない度合いを評価することができる。

【0074】

また、問いと回答を繰り返し音声入力し、各回答について評価点を算出した場合には、 図3のステップ S_{b14} 、 S_{b15} 、 S_{b16} においては、各回答について算出した評価点を加算するようにしてもよい。

【0075】

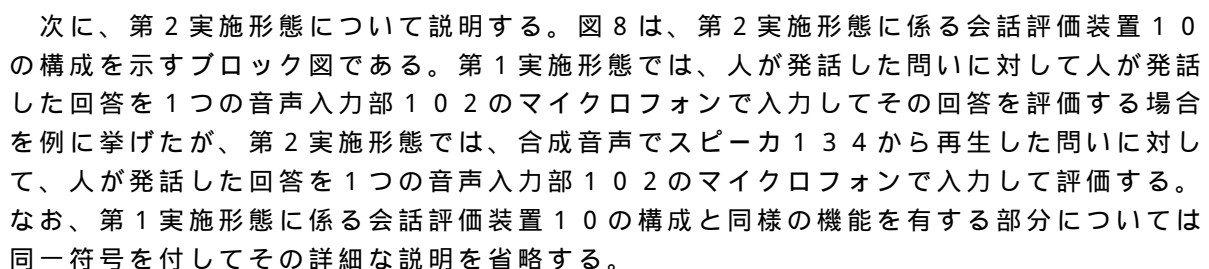
以上詳述したように、本実施形態に係る会話評価装置 10 によれば、問いに対する回答の音声特徴を問いの音声特徴との比較で評価することができる。これにより、その問いに対する回答として相手に与える印象を客観的に確認することができる。また、問いと回答の音声特徴として、問いの音高と回答の音高とは、相手に与える印象に密接な関係があるので、回答の音高を問いの音高との比較で評価することで、問いに対する回答について信頼性の高い評価をすることができる。さらに、問いと回答の音声特徴として、音高の他にも、問いの終了から回答の開始までの時間（会話間隔）は相手に与える印象に密接な関係がある。このため、問いと回答の音高のみならず、問いと回答の会話間隔についても評価することで、問いに対する回答についてより信頼性の高い評価をすることができる。

【0076】

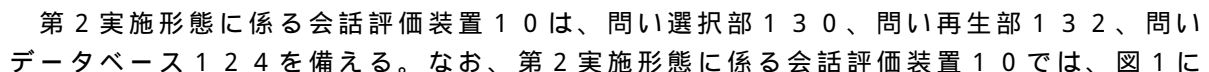
なお、第1実施形態にかかる会話評価装置 10 をスマートフォンや携帯電話機のような端末装置に適用した場合には、音声の入力と特徴の取得は携帯端末で行い、会話の評価については携帯端末とネットワークで接続された外部サーバが行うようにしてもよい。また、音声の入力は携帯端末で行い、入力した音声の特徴の取得と会話の評価については外部サーバが行うようにしてもよい。

【0077】

< 第2実施形態 >

次に、第2実施形態について説明する。 図8は、第2実施形態に係る会話評価装置 10 の構成を示すブロック図である。第1実施形態では、人が発話した問いに対して人が発話した回答を1つの音声入力部 102 のマイクロフォンで入力してその回答を評価する場合を例に挙げたが、第2実施形態では、合成音声でスピーカ 134 から再生した問いに対して、人が発話した回答を1つの音声入力部 102 のマイクロフォンで入力して評価する。なお、第1実施形態に係る会話評価装置 10 の構成と同様の機能を有する部分については同一符号を付してその詳細な説明を省略する。

【0078】

第2実施形態に係る会話評価装置 10 は、問い選択部 130、問い再生部 132、問いデータベース 124 を備える。なお、第2実施形態に係る会話評価装置 10 では、 図1に

10

20

30

40

50

示す判別部 108、言語データベース 122 が設けられていない。これは、第 2 実施形態に係る会話評価装置 10 では、問いは予め音高が決められている音声データが選択され、スピーカ 134 から再生されるので、発話が問いである否かの判定は不要だからである。

【0079】

問いデータベース 124 は、問いの音声データを、予め複数記憶する。この音声データは、モデルとなる人物の音声を録音したものである。問いの音声データについては、例えば wav や mp3 などのフォーマットであり、標準で再生したときの波形サンプル毎（または波形周期毎）の音高と、特定区間の音高（語尾区間の最高音高）が予め求められていて、その特定区間の音高を示すデータが音声データに対応付けられて問いデータベース 124 に記憶されている。なお、ここでいう標準で再生とは、音声データを録音時の条件（音高・音量・音色・話速など）と同じ条件で再生する、という意味である。

10

【0080】

なお、問いデータベース 124 に記憶する問いの音声データについては、人物 A、B、C、... のように複数人にわたって、同一内容の問いを記憶させても良い。人物 A、B、C、... については例えば有名人、タレント、歌手などとして、各人物毎に音声データをデータベース化する。また、このようにデータベース化する場合、メモリーカードなどの媒体を介して問いの音声データを問いデータベース 124 に格納させても良いし、会話評価装置 10 にネットワーク接続機能を持たせて、特定のサーバから問いの音声データをダウンロードし、問いデータベース 124 に格納させても良い。メモリーカードやサーバから問いの音声データを入手する場合、無償であっても良いし、有償であっても良い。

20

【0081】

また、問いの音声データは、どの人物をモデルとして欲しいのかを、利用者が操作入力部等によって選択可能な構成としても良いし、各種条件（日、週、月など）毎にランダムで決定する構成としても良い。また、問いの音声データは、音声入力部 102 のマイクロフォンを介して、利用者自身や、当該利用者の家族、知人の音声を録音したもの（または別途の装置によってデータ化したもの）をデータベース化しても良い。このように身近な人物の音声で問いが発話されると、あたかも当該人物と対話しているかのような感覚を得ることができる。

【0082】

問い選択部 130 は、問いの音声データを、問いデータベース 124 から 1 つを選択し、当該選択した問いの音声データを、対応付けられた音高データとともに、読み出して取得する。問い選択部 130 は、取得した音声データは問い再生部 132 に供給し、音高データは解析部 106 に供給する。なお、問い選択部 130 が、複数の音声データのうち、1 つの音声データをどのようなルールで選択するかについては、例えばランダムでも良いし、図示しない操作部から選ぶようにしても良い。問い再生部 132 は、問い選択部 130 からの問いの音声データをスピーカ 134 で再生する。

30

【0083】

次に、このような第 2 実施形態に係る会話評価装置 10 の動作について説明する。図 9 は、第 2 実施形態に係る会話評価装置 10 における処理動作を示すフローチャートである。まず、ステップ Sc11 において、問い選択部 130 は問いデータベース 124 から問いを選択する。続いて、ステップ Sc12 において、問い選択部 130 は、選択した問いの音声データと特徴データ（音高データ）を取得する。問い選択部 130 は、取得した音声データを問い再生部 132 に供給し、音高データは解析部 106 に供給する。解析部 106 の第 1 音高取得部 106A は、問い選択部 130 からの問いの音高データを取得し、評価部 110 に供給する。

40

【0084】

続いて、ステップ Sc13 において、問い再生部 132 は、選択された問いの音声データをスピーカ 134 で再生する。そして、ステップ Sc14 において、問いの再生が終了したか否かを判断する。ステップ Sc14 において、問いの再生が終了したと判断すると、ステップ Sc15 にて会話間隔の計時を開始する。以降は、回答の発話の処理（ステッ

50

プSc16～Sc20)であり、図2における回答の発話の処理(ステップSa17～Sa21)と同様である。

【0085】

このような第2実施形態に係る会話評価装置10によれば、スピーカ134で問いが再生され、その問いに対する回答の音声を音声入力部102のマイクロフォンで入力すると、その回答の評価値が表示部112に表示される。これによれば、問いがスピーカ134で再生されるので、問いを発話する相手がいなくても、1人で問いに対する回答を訓練することができる。また、問いがスピーカ134で再生されるので、回答だけを音声入力部102のマイクロフォンで入力すれば足りるため、音声入力部102から入力される発話が問いか否かの判別が不要になる。

10

【0086】

なお、本実施形態における解析部106において、第1音高取得部106Aは、音声入力部102を介さずに、問い選択部130により選択された問いの音声データを解析して、当該音声データを標準で再生したときの平均音高を取得し、この音高データを評価部110に供給する構成としても良い。この構成によれば、音高データを問いの音声データに予め関連付けて問いデータベース124に記憶させる必要がなくなる。

【0087】

<第3実施形態>

次に、第3実施形態について説明する。図10は、第3実施形態に係る会話評価装置10の構成を示すブロック図である。第1実施形態では、2人の会話音声を1つの音声入力部102のマイクロフォンで入力する場合を例に挙げたが、第3実施形態では、2人の会話音声を2つの音声入力部102A、102Bのそれぞれのマイクロフォンで別々に入力する。なお、第1実施形態に係る会話評価装置10の構成と同様の機能を有する部分については同一符号を付してその詳細な説明を省略する。

20

【0088】

第3実施形態に係る会話評価装置10では、図1に示す判別部108、言語データベース122が設けられていない。これは、第3実施形態に係る会話評価装置10では、各人の音声を別々の音声入力部102A、102Bで入力するので、問いを発する人と回答をする人を決めれば、発話が問いである否かの判定は不要だからである。

【0089】

次に、このような第3実施形態に係る会話評価装置10の動作について説明する。図11は、第3実施形態に係る会話評価装置10における処理動作を示すフローチャートである。図11に示すフローチャートは、図2に示すフローチャートの発話が問いか否かの判断処理をなくしたものである。さらに図11に示すステップSd11、Sd12、Sd13は、図2に示すステップSa11、Sa12、Sa15において「発話」とあるのを「問い」としたものである。以降の図11に示すステップSd14～Sd19は、図2に示すステップSa16～Sa21と同様である。

30

【0090】

このような第3実施形態に係る会話評価装置10によれば、例えば問いの音声が入力部102Aのマイクロフォンで入力すると、その回答の音声は別の音声入力部102Bのマイクロフォンで入力される。これにより、回答の評価値が表示部112に表示される。これによれば、問いと回答が音声入力部102A、102Bのそれぞれのマイクロフォンから別々に入力されるので、各音声入力部102A、102Bから入力される発話が問いか否かの判別が不要になる。

40

【符号の説明】

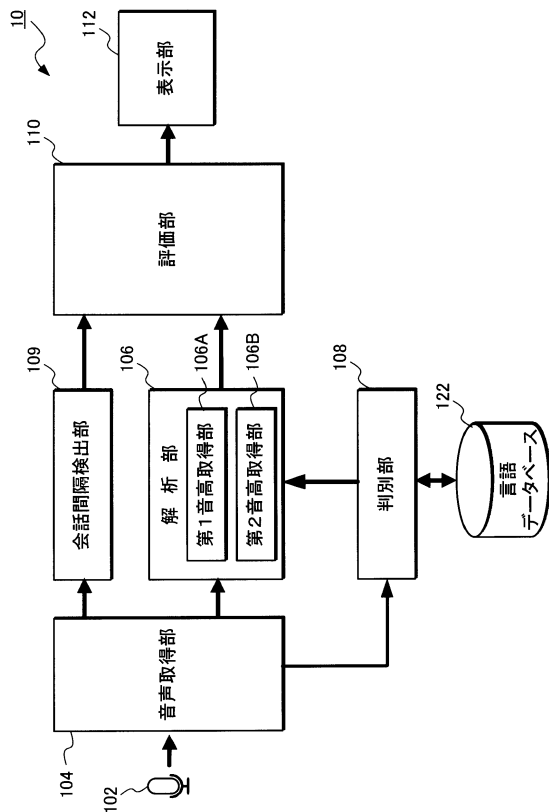
【0091】

10...会話評価装置、102(102A、102B)...音声入力部、104...音声取得部、106...解析部、106A...第1音高取得部、106B...第2音高取得部、108...判別部、109...会話間隔検出部、110...評価部、112...表示部、122...言語データベース、124...問いデータベース、130...選択部、132...問い再生部、134...ス

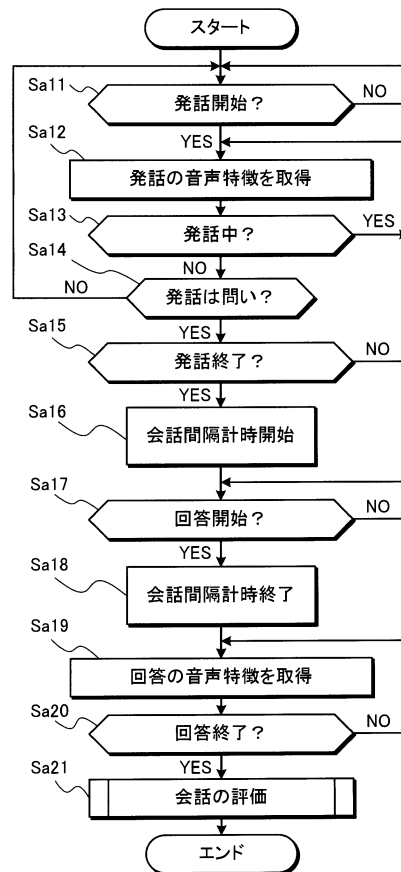
50

ピーカ。

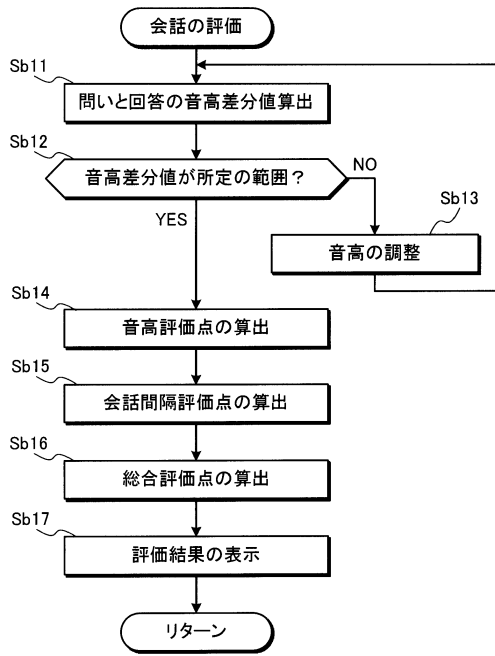
【図1】



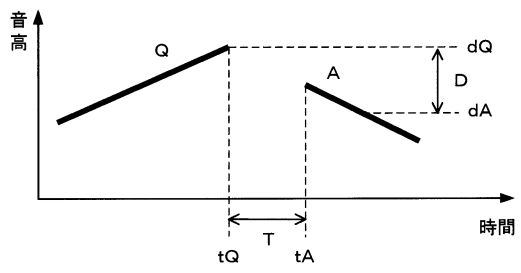
【図2】



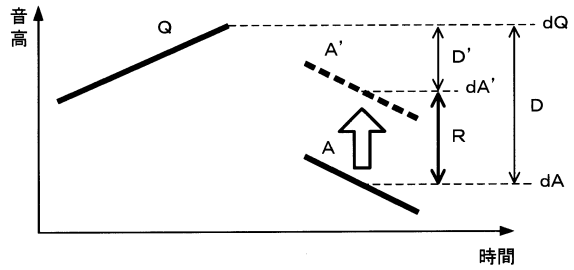
【図3】



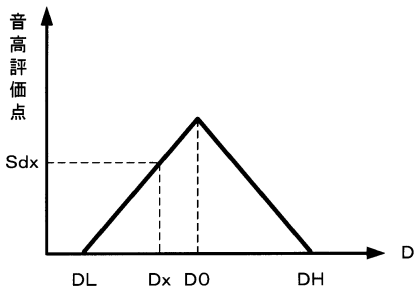
【図4】



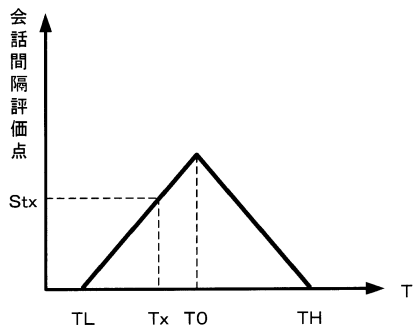
【図5】



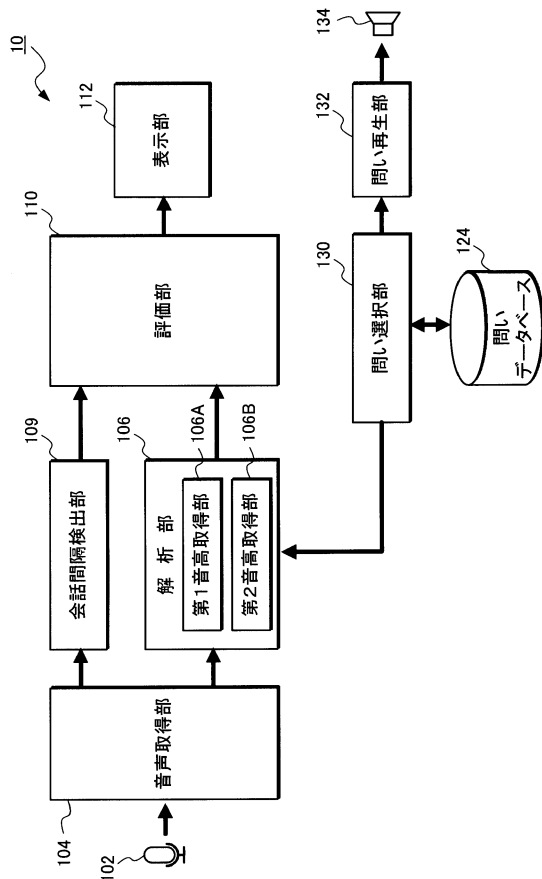
【図6】



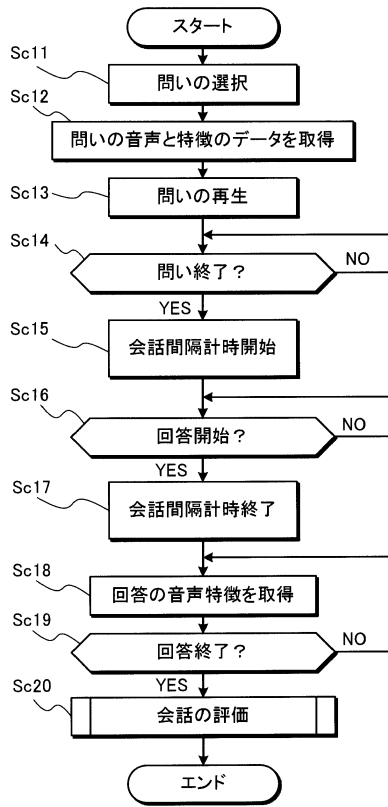
【図7】



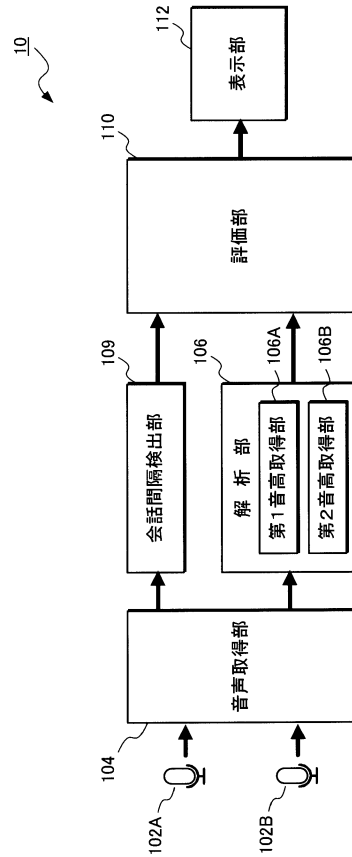
【図8】



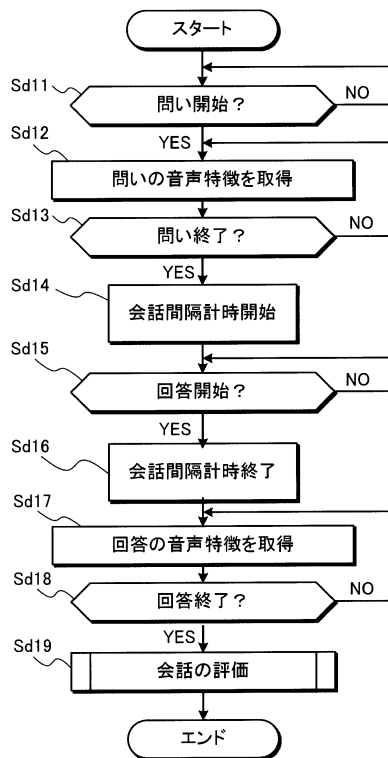
【図9】



【図10】



【図11】



フロントページの続き

(56)参考文献 特表2004-514178(JP,A)
特開2010-054568(JP,A)

(58)調査した分野(Int.Cl., DB名)

G10L 25/63

G10L 15/10

G10L 25/60

G10L 25/90