

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5534517号
(P5534517)

(45) 発行日 平成26年7月2日(2014.7.2)

(24) 登録日 平成26年5月9日(2014.5.9)

(51) Int. Cl.		F 1
G09B 19/04	(2006.01)	G09B 19/04
G09B 19/06	(2006.01)	G09B 19/06

請求項の数 4 (全 24 頁)

(21) 出願番号	特願2010-190481 (P2010-190481)	(73) 特許権者	000125369
(22) 出願日	平成22年8月27日 (2010.8.27)		学校法人東海大学
(65) 公開番号	特開2012-47998 (P2012-47998A)		東京都渋谷区富ヶ谷2丁目28番4号
(43) 公開日	平成24年3月8日 (2012.3.8)	(74) 代理人	100064414
審査請求日	平成25年8月22日 (2013.8.22)		弁理士 磯野 道造
		(74) 代理人	100111545
			弁理士 多田 悦夫
		(72) 発明者	山田 光穂
			東京都港区高輪2-3-23 東海大学内
		(72) 発明者	尾上 拓
			神奈川県平塚市北金目四丁目1番1号 東海大学内
		(72) 発明者	笠原 篤之
			神奈川県平塚市北金目四丁目1番1号 東海大学内

最終頁に続く

(54) 【発明の名称】 発話学習支援装置およびそのプログラム

(57) 【特許請求の範囲】

【請求項1】

撮影手段で撮影された、発話者がある言語の予め指定された単語を発話している際の口唇部分の画像から、当該発話者の口唇動作を得て、この口唇動作と、指導者が前記単語を発話したときの模範的な口唇動作とを比較した結果に基づいて、発話者の前記口唇動作の改善点を示す発話学習支援装置であって、

少なくとも発話者に発話させる前記単語と、指導者が前記単語を発話した際の口唇部分の画像との対応付けを複数記憶するデータ記憶手段と、

前記データ記憶手段に記憶された複数の前記単語の中から、発話者に発話させる単語を、外部からの入力により、あるいは、予め設定された順序により一つ指定し、当該単語の発話を発話者に指示する発話内容指定手段と、

前記発話内容指定手段で指定された前記単語を発話者が発話した際の、前記口唇部分の画像から、発話者の口唇動作を特定する基準となる予め設定した特徴点の位置を複数抽出する画像処理手段と、

前記画像処理手段で抽出された前記特徴点ごとの位置の変化を、口唇動作の履歴である動作履歴として測定する動作測定手段と、

前記動作測定手段で測定された前記特徴点ごとの前記動作履歴を数値解析することで、前記特徴点ごとに、予め設定した複数のスペクトル成分で表される動作スペクトルに変換するデータ変換手段と、

前記データ変換手段で求められた前記特徴点ごとの前記動作スペクトルと、前記模範的

10

20

な口唇部分の画像から予め求めた前記特徴点ごとの模範的な前記動作スペクトルとの差分を算出する差分算出手段と、

前記差分算出手段で算出された前記差分の絶対値と、予め定めた閾値とを前記特徴点ごとと比較し、前記差分の絶対値が前記予め定めた閾値よりも大きい前記特徴点がある場合、予め定めた補正関数によって、当該特徴点の動作を修正する方向と大きさを特定した修正量を算出する修正量算出手段と、

前記修正量算出手段で算出された前記修正量に応じた修正情報を表示装置に出力する修正情報出力手段と、を備えることを特徴とする発話学習支援装置。

【請求項 2】

前記修正情報出力手段は、

発話者の口唇部分の前記画像上における、前記修正量算出手段によって前記修正量が算出された前記特徴点に対応する位置に、当該特徴点の動作を修正する方向と大きさを示す画像を合成して前記表示装置に出力することを特徴とする請求項 1 に記載の発話学習支援装置。

【請求項 3】

前記修正情報出力手段は、

前記修正量算出手段によって前記修正量が算出された前記特徴点と当該特徴点の動作を修正する方向と大きさを特定したテキストを前記表示装置に出力することを特徴とする請求項 1 または請求項 2 に記載の発話学習支援装置。

【請求項 4】

撮影手段で撮影された、発話者がある言語の予め指定された単語を発話している際の口唇部分の画像から、当該発話者の口唇動作を得て、この口唇動作と、指導者が前記単語を発話したときの模範的な口唇動作とを比較した結果に基づいて、発話者の前記口唇動作の改善点を示すために、コンピュータを、

少なくとも発話者に発話させる前記単語と、指導者が前記単語を発話した際の口唇部分の画像との対応付けを複数記憶するデータ記憶手段に記憶された複数の前記単語の中から、発話者に発話させる単語を、外部からの入力により、あるいは、予め設定された順序により一つ指定し、当該単語の発話を発話者に指示する発話内容指定手段、

前記発話内容指定手段で指定された前記単語を発話者が発話した際の、前記口唇部分の画像から、発話者の口唇動作を特定する基準となる予め設定した特徴点の位置を複数抽出する画像処理手段、

前記画像処理手段で抽出された前記特徴点ごとの位置の変化を、口唇動作の履歴である動作履歴として測定する動作測定手段、

前記動作測定手段で測定された前記特徴点ごとの前記動作履歴を数値解析することで、前記特徴点ごとに、予め設定した複数のスペクトル成分で表される動作スペクトルに変換するデータ変換手段、

前記データ変換手段で求められた前記特徴点ごとの前記動作スペクトルと、前記模範的な口唇部分の画像から予め求めた前記特徴点ごとの模範的な前記動作スペクトルとの差分を算出する差分算出手段、

前記差分算出手段で算出された前記差分の絶対値と、予め定めた閾値とを前記特徴点ごとと比較し、前記差分の絶対値が前記予め定めた閾値よりも大きい前記特徴点がある場合、予め定めた補正関数によって、当該特徴点の動作を修正する方向と大きさを特定した修正量を算出する修正量算出手段、

前記修正量算出手段で算出された前記修正量に応じた修正情報を表示装置に出力する修正情報出力手段、として機能させるための発話学習支援プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、主に語学の発話学習等に利用されるものであり、発話者の発話時の口唇動作を、模範的な口唇動作と比較し、発話者の口唇動作の改善点を指示する発話学習支援装置

10

20

30

40

50

およびそのプログラムに関する。

【背景技術】

【0002】

従来、日本では、語学の発話学習において、正しい発音を習得するために、発話者が教科書に書かれた母音ごとの口唇や舌の動きを模擬し、同時に指導者の発音を耳で聞きながら自習する、もしくは、指導者の前で発音を繰り返すことで指導者の発音に近づけるという学習方法が広く行われてきた。

また、聾話者が発音を学習する場合、自分の声が聞こえないため、指導者の前で口唇や舌の動きを模擬して発音を繰り返し、指導者がその発音を聞いて改善点をアドバイスすることで、指導者の発音に近づけるという学習方法が広く行われてきた。

10

【0003】

しかし、前記した学習方法によると、発話者は、自分の発音が正しいかどうかを確認したい場合、テープに発音時の声を録音して後で聞くことで、発話者自身が判断することになるが、発話者が自身の習熟度を正確に判断することが困難であった。また、聾話者の場合、自身の声を聞くことができないため、口唇の動きをどのように変えれば発音が良くなるのかを客観的に判断することができなかった。

【0004】

これに鑑み、正しい発音を習得するための学習支援システムとして、発話者に単語を発音させた際の音声データと、指導者が同じ単語を発音した際の音声データとを比較することで、発話者の発音を採点あるいは評価したり、改善点を指示したりするものが開発されている。

20

【0005】

例えば、特許文献1には、発話者に語学学習の授業を提供する学習サーバ装置であって、PC(Personal Computer)から送信される発話者の語学学習の音声を解析して発話者の発音を採点する技術が開示されている。

また例えば、特許文献2には、ユーザの音声を取得して、予め記憶された模範音声と、取得したユーザの音声を比較し、その比較結果に基づいて、模範音声とユーザの音声との相違点を抽出し、抽出された相違点が存在する部分について強調すべき態様を指示する強調指示データを生成し、生成された強調指示データに基づく態様に合わせて模範音声を出力する語学学習装置が開示されている。

30

さらに例えば、特許文献3には、音韻毎に連結したデータである指導者データを1以上格納しておき、発話者の音声の入力を受け付けると、この音声をフレームに区分し、フレーム毎の音声データを1以上取得し、指導者データと1以上のフレーム毎の音声データに基づいて、発話者の音声の評定を行い、評定結果を出力する発音評定装置が開示されている。

【先行技術文献】

【特許文献】

【0006】

【特許文献1】特開2007-212558号公報

【特許文献2】特開2007-139868号公報

【特許文献3】特開2006-227587号公報

40

【発明の概要】

【発明が解決しようとする課題】

【0007】

一方で、英米の英語圏の国では、語学学習において、正しい発音を習得するためには、正しい口唇動作を習得することが重要であると考えられてきた。

このため、英米の英語圏の国では、指導者が単語を発音した際の口唇動作を発話者に見せて、指導者の口唇動作を発話者に真似させ、そのときの口唇動作を、指導者の口唇動作と比較して評価したり、改善点をアドバイスしたりすることで、発話者の口唇動作を指導者の口唇動作に近づけさせることが、発話者に正しい発音を習得させるための有効な指導

50

法として確立している。

【0008】

しかしながら、特許文献1～3に記載の従来の発話学習支援システムでは、発話者が発話した際の音声解析して発話者の発音を評価するものであるため、発話者の口唇動作の改善点を指示することができなかった。このため、発話者が、自身の口唇動作の改善点を客観的に認識できるようにする技術の確立が望まれていた。

【0009】

本発明は、前記した従来技術の問題を解決するために成されたもので、発話者の口唇動作を指導者の口唇動作と比較した結果に基づいて、発話者の口唇動作の改善点を客観的に示すことが可能な発話学習支援装置およびそのプログラムを提供することを目的とする。

10

【課題を解決するための手段】

【0010】

前記課題を解決するため、請求項1に記載の発明は、撮影手段で撮影された、発話者がある言語の予め指定された単語を発話している際の口唇部分の画像から、当該発話者の口唇動作を得て、この口唇動作と、指導者が前記単語を発話した際の模範的な口唇動作とを比較した結果に基づいて、発話者の前記口唇動作の改善点を示す発話学習支援装置であって、データ記憶手段と、発話内容指定手段と、画像処理手段と、動作測定手段と、データ変換手段と、差分算出手段と、修正量算出手段と、修正情報出力手段と、を備える構成とした。

【0011】

20

かかる構成によれば、発話学習支援装置は、データ記憶手段によって、少なくとも発話者に発話させる前記単語と、指導者が前記単語を発話した際の口唇部分の画像との対応付けを複数記憶する。

また、発話学習支援装置は、発話内容指定手段によって、前記データ記憶手段に記憶された複数の前記単語の中から、発話者に発話させる単語を、外部からの入力により、あるいは、予め設定された順序により一つ指定し、当該単語の発話を発話者に指示する。発話内容指定手段は、例えば、発話者に発話させる単語をデータ記憶手段から読み出し、この単語を表示装置に表示することにより、発話者に当該単語の発話を指示する。

【0012】

また、発話学習支援装置は、画像処理手段によって、前記発話内容指定手段で指定された前記単語を発話者が発話した際の、前記口唇部分の画像から、発話者の口唇動作を特定する基準となる予め設定した特徴点の位置を複数抽出する。

30

【0013】

さらに、発話学習支援装置は、動作測定手段によって、前記画像処理手段で抽出された前記特徴点ごとの位置の変化を、口唇動作の履歴である動作履歴として測定する。

そして、発話学習支援装置は、データ変換手段によって、前記動作測定手段で測定された前記特徴点ごとの前記動作履歴を数値解析することで、前記特徴点ごとに、予め設定した複数のスペクトル成分で表される動作スペクトルに変換する。

【0014】

そして、発話学習支援装置は、差分算出手段によって、前記データ変換手段で求められた前記特徴点ごとの前記動作スペクトルと、前記模範的な口唇部分の画像から予め求めた前記特徴点ごとの模範的な前記動作スペクトルとの差分を算出する。

40

【0015】

さらに、発話学習支援装置は、修正量算出手段によって、前記差分算出手段で算出された前記差分の絶対値と、予め定めた閾値とを前記特徴点ごとに比較し、前記差分の絶対値が前記予め定めた閾値よりも大きい前記特徴点がある場合、予め定めた補正関数によって、当該特徴点の動作を修正する方向と大きさを特定した修正量を算出する。

そして、発話学習支援装置は、修正情報出力手段によって、前記修正量算出手段で算出された前記修正量に応じた修正情報を表示装置に出力する。

【0016】

50

これによれば、発話者の口唇動作を指導者の口唇動作と比較して、発話者の口唇動作の修正量を求め、その修正量に応じた修正情報を表示装置に表示することで、発話者に、口唇動作の改善点を客観的に認識させることができる。

【 0 0 1 7 】

また、請求項 2 に記載の発話学習支援装置は、請求項 1 に記載の発話学習支援装置において、前記修正情報出力手段は、発話者の口唇部分の前記画像上における、前記修正量算出手段によって前記修正量が算出された前記特徴点に対応する位置に、当該特徴点の動作を修正する方向と大きさを示す画像を合成して前記表示装置に出力することを特徴とする。

【 0 0 1 8 】

例えば、発話者の口唇部分の画像上における、動作を修正する特徴点に対応する位置に、修正すべき方向と大きさを特定する図形（例えば矢印）の C G（Computer Graphics）を合成して表示することができる。この図形の C G は、予め適宜の記憶手段に記憶されていてもよいし、修正量算出手段によって算出された修正量に基づいて、修正情報出力手段が、その都度生成してもよい。

これによれば、発話者に自身の口唇動作をどのように修正すればよいのかを、直感的に認識させやすくなる。

【 0 0 1 9 】

また、請求項 3 に記載の発話学習支援装置は、請求項 1 または請求項 2 に記載の発話学習支援装置において、前記修正情報出力手段は、前記修正量算出手段によって前記修正量が算出された前記特徴点と当該特徴点の動作を修正する方向と大きさを特定したテキストを前記表示装置に出力することを特徴とする。

例えば、当該テキストを音声合成し、表示装置のスピーカで再生してもよいし、当該テキストを、表示装置の表示画面に表示してもよい。

これによれば、どの特徴点の動作をどの程度修正すればよいのかがテキストで表示されるので、発話者に自身の口唇動作の改善点をより理解させやすくなる。

【 0 0 2 0 】

また、請求項 4 に記載の発話学習支援プログラムは、撮影手段で撮影された、発話者がある言語の予め指定された単語を発話している際の口唇部分の画像から、当該発話者の口唇動作を得て、この口唇動作と、指導者が前記単語を発話した際の模範的な口唇動作とを比較した結果に基づいて、発話者の前記口唇動作の改善点を示すために、コンピュータを、発話指示手段、画像処理手段、動作測定手段、データ変換手段、差分算出手段、修正量算出手段、修正情報出力手段、として機能させることを特徴とする。

【 0 0 2 1 】

かかる構成によれば、発話学習支援プログラムは、発話指示手段によって、少なくとも発話者に発話させる前記単語と、指導者が前記単語を発話した際の口唇部分の画像との対応付けを複数記憶するデータ記憶手段に記憶された複数の前記単語の中から、発話者に発話させる単語を外部からの入力により、あるいは、予め設定された順序により一つ選択し、発話者に対し当該単語の発話を指示する。

【 0 0 2 2 】

発話学習支援プログラムは、画像処理手段によって、前記発話指示手段で指示された前記単語を発話者が発話した際の、前記口唇部分の画像から、発話者の口唇動作を特定する基準となる予め設定した特徴点の位置を複数抽出する。

【 0 0 2 3 】

発話学習支援プログラムは、動作測定手段によって、前記画像処理手段で抽出された前記特徴点ごとの位置の変化を、口唇動作の履歴である動作履歴として測定する。

【 0 0 2 4 】

発話学習支援プログラムは、データ変換手段によって、前記動作測定手段で測定された前記特徴点ごとの前記動作履歴を数値解析することで、前記特徴点ごとに、予め設定した複数のスペクトル成分で表される動作スペクトルに変換する。

10

20

30

40

50

【 0 0 2 5 】

発話学習支援プログラムは、差分算出手段によって、前記動作測定手段で測定された前記特徴点ごとの前記動作履歴を数値解析することで、前記特徴点ごとに、予め設定した複数のスペクトル成分で表される動作スペクトルに変換する。

【 0 0 2 6 】

発話学習支援プログラムは、修正量算出手段によって、前記差分算出手段で算出された前記差分の絶対値と、予め定めた閾値とを前記特徴点ごとに比較し、前記差分の絶対値が前記予め定めた閾値よりも大きい前記特徴点がある場合、予め定めた補正関数によって、当該特徴点の動作を修正する方向と大きさを特定した修正量を算出する。

【 0 0 2 7 】

そして、発話学習支援プログラムは、修正情報出力手段によって、前記修正量算出手段で算出された前記修正量に応じた修正情報を表示装置に出力する。

【 発明の効果 】

【 0 0 2 8 】

本発明に係る発話学習支援装置及び発話学習支援プログラムでは、以下のような優れた効果を奏する。

請求項 1、4 に記載の発明によれば、発話者の口唇動作を模範的な口唇動作と比較した結果に基づいて、修正量を算出し、修正量に対応する修正情報を表示装置に表示させることで、発話者に対し、自身の口唇動作の改善点を客観的に示すことができるので、正しい発音を効果的に学習可能となる。

請求項 2、3 に記載の発明によれば、発話者に、口唇動作の修正点をより理解させやすくなる。

【 図面の簡単な説明 】

【 0 0 2 9 】

【 図 1 】本発明における発話学習支援装置を用いて発話者が学習する様子を概念的に説明するための図であり、学習開始前に表示画面に表示する画面例と、学習開始後に表示画面に表示する画面例を示している。

【 図 2 】本発明の発話学習支援装置の構成を示すブロック構成図である。

【 図 3 】本発明の発話学習支援装置におけるデータ記憶手段が記憶するデータ構造の例を概念的に表した図である。

【 図 4 】本発明の発話学習支援装置における画像処理手段により抽出する特徴点を示した図である。

【 図 5 】本発明の発話学習支援装置における動作測定手段により生成される動作履歴グラフの例を示した図である。

【 図 6 】本発明の発話学習支援装置におけるデータ変換手段により生成される特徴点の動作スペクトルグラフの例を示した図である。

【 図 7 】英単語の中で日本語の母音に近い母音列を有する単語を発話した際の特徴点の動作スペクトルグラフの例を示した図である。

【 図 8 】英単語の中で日本語の母音に近い母音列を有する英母音間の特徴点の動作スペクトルグラフの比較例を示した図である。

【 図 9 】(a) は、指導者がある英母音を含む単語を発話した際の特徴点の動作スペクトルにおけるスペクトル成分の比率を示し、(b) は、発話者(学習者)が、学習を始めた当初に、(a) と同じ英母音を含む単語を発話した際の特徴点の動作スペクトルにおけるスペクトル成分の比率を示し、(c) は、(b) と同じ発話者(学習者)が、学習がある程度進んだ段階で、(a) および(b) と同じ英母音を含む単語を発話した際の特徴点の動作スペクトルにおけるスペクトル成分の比率を示す図である。

【 図 1 0 】本発明の発話学習支援装置における修正量算出手段で用いられる補正関数を説明するための概念図である。

【 図 1 1 】本発明の発話学習支援装置における修正情報出力手段により発話者の口唇動作の修正点を表示した際の画面例を示す図である。

10

20

30

40

50

【図 1 2】本発明の発話学習支援装置の動作を示すフローチャートである。

【図 1 3】本発明の実施形態の変形例に係る発話学習支援装置の構成を示すブロック構成図である。

【発明を実施するための形態】

【 0 0 3 0 】

以下、本発明の実施の形態について図面を参照して説明する。

[発話学習支援装置の概要]

まず、図 1 を参照して、本発明における発話学習支援装置 1 の概要について説明する。

発話学習支援装置 1 は、発話者が学習しようとする言語の指定された単語を発話した際の口唇動作と、その言語の学習を指導する教師あるいはその言語を母国語とするネイティブスピーカー等（以下では、単に指導者という。）に、同じ単語を予め発話してもらった際の模範的な口唇動作と、を比較した結果に基づいて、発話者の口唇動作の改善点を示すものである。

10

【 0 0 3 1 】

図 1 右側に示すように、発話学習支援装置 1 は、発話者に発話させる単語を指示するテキストデータ T a と、その単語を指導者が発話した際の模範的な口唇部分の画像 T G とを、表示装置 3 0 の表示画面 3 1 に表示させる。この画像 T G は、発話学習支援装置 1 が予め保有するものであり、発話者に発話させる単語を指導者が発話した際の、指導者の口唇部分の一連の動作を示す動画像である。

【 0 0 3 2 】

また、発話学習支援装置 1 は、図 1 左側に示すように、初期画面として、例えば言語選択ボタン、難易度選択ボタン、学習開始ボタン等を備えた画面を表示画面 3 1 に表示するようになっている。この初期画面は、発話学習支援装置 1 の図示しない記憶手段に予め記憶されており、発話者が発話学習支援システムを起動させた際に、表示装置 3 0 の表示画面 3 1 に表示されるようになっている。一方、表示装置 3 0 は、例えばマウス等を介して、発話者による学習する言語の選択、難易度の選択、学習開始の決定の入力を受け付けた際は、そのことを示す信号を発話学習支援装置 1 に出力するようになっている。なお、この初期画面において、発話者によって、言語選択ボタン等で学習する言語の選択が入力された後に、母音選択画面に遷移するようになっていてもよい。これによれば、発話者が苦手な母音を重点的に学習することができるので、利便性が高くなる。

20

30

【 0 0 3 3 】

そして、発話学習支援装置 1 は、表示装置 3 0 から、例えば、発話者により学習言語として「英語」が選択され、難易度として「普通」が選択されたことを示す信号の入力を受け付けた場合、図 1 右側に例示するように、発話者に発話させる単語を指示する“「 a u n t 」を発話してください”というテキストと、予め保有する、指導者が「 a u n t 」を発話したときの口唇部分の画像 G t とを、表示画面 3 1 に表示する。

【 0 0 3 4 】

そして、発話者が、発話学習支援装置 1 によって指定されて表示画面 3 1 に表示された単語を、表示画面 3 1 に表示された模範的な口唇動作の画像 T G を参照しつつ発話すると、撮影手段 2 0 により、発話者の口唇部分が発話開始から発話終了まで継続して撮影される。

40

【 0 0 3 5 】

そして、発話学習支援装置 1 は、撮影手段 2 0 によって撮影された発話者の口唇部分の画像を取得して解析し、その解析結果と、指導者が同じ単語を発話した際の口唇部分の画像 T G の解析結果とを比較した結果に基づいて、発話者の口唇動作の修正量を算出し、その修正量に応じた修正情報を表示装置 3 0 の表示画面 3 1 に表示するようになっている。

【 0 0 3 6 】

このように、発話学習支援装置 1 によれば、発話者の口唇動作の修正量を算出し、その修正量に応じた修正情報を、表示装置 3 0 の表示画面 3 1 に表示するため、発話者に対し、口唇動作の改善点を客観的に示すことが可能となる。

50

【0037】

なお動画は、発話者が模範的な口唇動作を繰り返し確認できるよう、所定回数繰り返して再生されるように予め設定しておいてもよいし、発話者によって外部から信号を入力された際に繰り返して再生するようにしてもよい。

【0038】

[発話学習支援装置の構成]

次に、図2を参照して、本発明における発話学習支援装置1の構成について説明する。

図2に示すように、発話学習支援システムは、発話学習支援装置1と、撮影手段20と、表示装置30と、を含んで構成されている。

ここで、発話学習支援装置1の構成の説明に先立ち、適宜図1を参照して撮影手段20および表示装置30について説明する。

【0039】

撮影手段20は、発話者が発話している際に、当該発話者の口唇部分を撮影するためのものである。この撮影手段20で撮影された口唇部分の画像は、発話学習支援装置1に出力される。撮影手段20は、例えば、一般的なカメラであってもよいし、口唇部分の奥行きの変位を検出可能なステレオカメラであってもよい。なお、口唇部分の画像は、発話者が発話している間、撮影手段20によって継続して撮影されたものである。また、口唇部分とは、口唇の周囲だけではなく、発話者の顔面下部(鼻下から下顎まで)を含んでいてもよい。また、後記する画像処理手段13で、発話者の顔面全体の画像を使って画像処理を行う場合には、撮影手段20は、発話者の顔面全体を撮影してもよい。この撮影手段20は、図1に示すように、表示装置30に内蔵されていてもよい。

【0040】

表示装置30は、発話学習支援装置1から単語の発話を指示するテキストデータ、および、その単語を指導者が発話した際の模範的な口唇動作の画像の入力を受け付けて発話者に表示するものであり、表示画面31を有している。例えば表示装置30を、PC(Personal Computer)とし、表示画面31を、PCのモニタ等とすることができる。なお、図1では、図示の都合上、表示画面31に、模範的な口唇動作の一場面が表示されているが、実際は、模範的な口唇動作を動画で表示するようになっている。表示装置30は、さらにスピーカ(図示せず)を備えていてもよい。

【0041】

なお、表示画面31は、画像を平面的に表示するものに限らず、立体的に表示するものを用いてもよい。表示画面31が、画像を立体的に表示可能な場合には、模範的な口唇動作の動画を立体的に表示するとよい。立体的に表示することにより、「ウ、u」、「オ、o」等の口唇を突き出して発音する母音を含む単語の学習に好適である。

【0042】

発話学習支援装置1は、撮影手段20で撮影された、発話者がある言語の予め指定された単語を発話している際の口唇部分の画像から、当該発話者の口唇動作を得て、この口唇動作と、予め保有する模範的な口唇動作とを比較した結果に基づいて、発話者の口唇動作の改善点を示すものである。この発話学習支援装置1は、データ記憶手段11と、発話内容指定手段12と、画像処理手段13と、動作測定手段14と、データ変換手段15と、差分算出手段16と、修正量算出手段17と、修正情報出力手段18と、を備えている。なおここでは、発話学習支援装置1は、発話者が発話した際の音声を入力しておらず、撮影手段20で撮影された動画のみから発話者の口唇動作を得ている。

【0043】

データ記憶手段11は、少なくとも発話者に発話させる単語と、指導者がその単語を発話した際の口唇部分の画像との対応付けを複数記憶するものであり、不揮発性メモリ(NVRAM)、ハードディスク等の一般的な記憶媒体である。

【0044】

ここで、図3を参照しつつ、データ記憶手段11内のデータ構造の例を説明する。なお、図3では、データ記憶手段11内のデータ構造の一部を概念的に表している。

図3に示すように、データ記憶手段11は、ここでは、各言語に存在する単語と、発話者にその単語の発話を指示するテキストデータ(テキスト1, テキスト2, ..., テキストn)と、指導者がその単語を発話した際の口唇部分の画像(画像1, 画像2, ..., 画像n)と、指導者の口唇部分の画像から求めた特徴点ごとの模範的な動作スペクトル(スペクトル1, スペクトル2, ..., スペクトルn)と、当該動作スペクトルの第1成分が動作スペクトル全体に占める比率(比率1, 比率2, ..., 比率n)と、単語の発音の難易度(難しい, 普通, 易しい)と、を一組としたデータ組(データ1, データ2, ..., データn)とを、その言語の母音ごとに分類して複数記憶している。動作スペクトルおよび比率については、詳しくは後記する。

【0045】

一つの母音につき単語が少なくとも一つ記憶されていればよいので、データ記憶手段11は、一つの母音につき少なくとも一つのデータ組を記憶していれば足りるが、図3に示すように、一つの母音につき複数のデータ組を記憶しておくことが好ましい。一つの母音につき複数のデータ組を記憶しておくこと、発話者が、一つの母音について、様々な単語を発話して学習することが可能となり、発話者の学習効率を向上させることができるためである。なお、図3では、一つの母音につき2つのデータ組を記憶した例を示したが、これに限られず、任意数とすることができる。

【0046】

例えば、言語が日本語の場合、母音は「ア、イ、ウ、エ、オ」の5音となるので、データ記憶手段11は、予め定めた日本語の単語を、母音「ア」を含む単語、母音「イ」を含む単語、・・・、母音「オ」を含む単語にそれぞれ分類して記憶している。また、データ記憶手段11は、このほかに、長母音、短母音、複合母音、半母音、弱母音等、その言語に存在する母音ごとにそれぞれデータ組を記憶している。

【0047】

なお、データ記憶手段11は、一つの母音内において、複数のデータ組を予めある順序(例えば発音の難易度順等)に従って並べて記憶していてもよい。なお、図3のデータ構造は一例であり、これに限られるものではなく、例えば難易度を記憶していなくてもよい。このようにしてデータ記憶手段11に記憶された複数のデータ組の各データは、発話内容指定手段12あるいは差分算出手段16によって適宜読み出される。

【0048】

発話内容指定手段12は、データ記憶手段11に記憶された複数のデータ組の単語の中から、発話者に発話させる単語を、外部からの入力により、あるいは、予め設定された順序により一つ指定し、当該単語の発話を発話者に指示するものである。

発話内容指定手段12は、ここでは、表示装置30の表示画面31に表示する表示内容を制御する機能も有している。発話内容指定手段12は、ここでは、図示しない記憶手段から初期画面を読み出して表示画面31に表示すると共に、表示装置30から、初期画面において発話者によって各種項目が選択されたことを示す信号および学習開始の決定を示す信号の入力を受け付けると、データ記憶手段11に母音ごとに記憶された複数のデータ組の中から、発話者に発話させる単語を一つ選択し、その単語の発話を指示するテキストデータと、その単語を指導者が予め発話した際の口唇部分の画像と、を読み出して表示画面31に表示する。

【0049】

発話者に発話させる単語を指導者が予め発話した際の口唇部分の画像は、指導者の模範的な口唇動作を示すものであり、この画像を表示画面31に表示することで、発話者が模範的な口唇動作を参照しながら発話できることから、発話者の学習効率を向上させることができる。

【0050】

発話内容指定手段12は、発話者に発話させる単語を指定するときは、データ記憶手段11に母音ごとに記憶された複数のデータ組から、単語をランダムに選択してよい。例えば、データ記憶手段11に図3に示したデータ構造が記憶されている場合、発話内容指定

10

20

30

40

50

手段 1 2 は、発話者に発話させる単語を、まず「a u n t」とし、次に「c u p」とする等、ランダムに選択することができる。

【 0 0 5 1 】

ただし、初期画面において、発話者により難易度が指定された場合、発話内容指定手段 1 2 は、データ記憶手段 1 1 に記憶された複数のデータ組内の難易度を参照し、発話者により指定された難易度に適応した単語を選択するようにする。また、初期画面において、発話者により母音が指定された場合、発話内容指定手段 1 2 は、データ記憶手段 1 1 において、発話者により指定された母音に対応付けて記憶された複数のデータ組から単語を適宜選択する。

【 0 0 5 2 】

画像処理手段 1 3 は、撮影手段 2 0 で撮影された口唇部分の動画像から、当該口唇部分における予め設定した特徴点の位置を抽出するものである。

ここで、図 4 を参照して特徴点について説明する。図 4 に示すように、特徴点は、口唇部分の上端（口唇上部の上端点 2 点（ a_1 、 a_2 ）の中点）A 点、下端 B 点、左端 C 点、右端 D 点の 4 点、または、下顎の頂点をさらに含む 5 点とすることができる。なお、前記した 4 点は、口唇部分を示す赤色の画素値を有する画素、顔面部分を示す肌色の画素値を有する画素との画素値の差が最大となる画素（境界となる画素）、つまり、口唇上部（上唇）及び口唇下部（下唇）の最端座標に位置する画素が対象となっている。さらに下顎の頂点を抽出するときは、顎部の肌色の画素値を有する画素と首部の肌色の画素値を有する画素との画素値の差（輝度の差）が最大となる画素で、且つ、最下端に位置する画素を抽出すればよい。

【 0 0 5 3 】

この画像処理手段 1 3 は、撮影手段 2 0 から口唇部分の動画像が継続して入力される限り、各画像について、抽出点を抽出し続ける。つまり、画像処理手段 1 3 は、1 つの画像について、4 点あるいは 5 点を抽出し、動画像を構成する複数の画像（通常、1 秒間に 30 フレーム）ごとに抽出していき、抽出した特徴点を、口唇部分の動画像を撮影していた時刻に沿った時系列データとして出力することとなる。画像処理手段 1 3 は、抽出した特徴点の位置を、動作測定手段 1 4 に出力する。

【 0 0 5 4 】

なおここでは、画像処理手段 1 3 で、4 点あるいは 5 点の特徴点を抽出することとしたが、これに限られず任意点数（6 点以上）の特徴点を抽出してもよい。

【 0 0 5 5 】

動作測定手段 1 4 は、画像処理手段 1 3 で抽出された特徴点の位置について、当該特徴点の位置の変化を、口唇動作の履歴である動作履歴として測定するものである。

そして、この動作測定手段 1 4 では、画像処理手段 1 3 で抽出された 4 点あるいは 5 点の特徴点それぞれについて、特徴点それぞれの変化を表す動作履歴を測定する。なお、画像処理手段 1 3 に口唇部分の動画像が継続して入力され、特徴点が抽出された際に、一定時間（例えば 2 秒間）、特徴点の変化が無い場合、つまり、口唇部分の動画像には変化がない（動きがない）場合、この動作測定手段 1 4 によって、発話者は発話していないと判定される。

【 0 0 5 6 】

ここで、図 5 を参照して、動作履歴（動作履歴グラフ）の例について説明する。この図 5 に示した動作履歴グラフは、口唇下部（下唇）の B 点（図 4 参照）の変化について示したものであり、横軸に時間 t （ms）、縦軸に変位 y （mm）を取ったものである。この動作履歴グラフの例に示したように、下唇は、時間経過により、一旦下がって（14 ms で最もさがる（約 5.1 mm））、その後上がっている。動作測定手段 1 4 によって測定された動作履歴（動作履歴グラフ）は、データ変換手段 1 5 に出力される。

【 0 0 5 7 】

図 2 に戻って、データ変換手段 1 5 は、動作測定手段 1 4 で測定された動作履歴（動作履歴グラフ）を、数値解析することで、予め設定した複数のスペクトル成分で表される動

10

20

30

40

50

作スペクトルグラフに変換するものである。データ変換手段 1 5 では、動作測定手段 1 4 で測定された 4 つあるいは 5 つの動作履歴それぞれを、動作スペクトルグラフに変換する。この実施形態では、数値解析にフーリエ変換を用いているが、フーリエ変換以外に、動作履歴を関数に見立てて近似する関数近似や動作履歴について微小時間ごとの特徴点の移動量を基に解析する方法等を採用してもよい。

【 0 0 5 8 】

ここで、図 6 を参照して、動作スペクトルグラフの例について説明する。図 6 は、横軸に周波数を取って、縦軸にスペクトル強度を取ったものである。

図 6 (a) は、発話者が「 Λ 」という母音を発話した際に、画像処理手段 1 3 で口唇上部の特徴点を抽出し、動作測定手段 1 4 で動作履歴を測定したものをデータ変換手段 1 5 で変換したものである。また、図 6 (b) は、(a) と同じ発話者が「 ε 」という母音を発話した際に、画像処理手段 1 3 で口唇上部の特徴点を抽出し、動作測定手段 1 4 で動作履歴を測定したものをデータ変換手段 1 5 で変換したものである。この二つの例を比較すると、発話者が母音「 Λ 」を発話したときに得られる動作スペクトルでは、動作スペクトルの第 1 成分が動作スペクトル全体のほとんどを占める一方で、同じ発話者が母音「 ε 」を発話したときに得られる動作スペクトルでは、動作スペクトルの第 1 成分に加え、動作スペクトルの第 2 成分以降の成分（特に第 2 成分）も多く含んでいる。このように、二つの母音「 Λ 」および「 ε 」間の動作スペクトルは大きく異なることが分かる。

【 0 0 5 9 】

このように、動作スペクトルグラフは、特徴的なピークを有したものの、この例では、検出されたスペクトルにおいて、周波数が“ 1 ”のところにピークが存在する。この特徴的なピークは、発話者が発話する際に、口唇部分の動作が以下に述べるようになるために生じるものである。すなわち、口唇部分の動作は、発話時にまず開かれ、ある形状に変形され、母音一語が発音された後に、閉じられるという一連の動作の繰り返しになる。そしてこのとき、母音一語の発音では、口唇部分が開かれてから閉じられるまでの一連の動作について、多少の個人差はあるものの、動作履歴グラフが非常に周期的になる。この結果、母音発音時には、必ず特徴的な周波数が現出することになる。

【 0 0 6 0 】

なお、データ変換手段 1 5 において、フーリエ変換を用いて動作スペクトルグラフを識別する場合、発話者が発話した際の動作速度（発話速度）が速くなったり遅くなったりしても、現出する様々なパターン（スペクトルパターン）は、周波数の位置が異なるだけで、パターン形状が変化することはない。つまり、このデータ変換手段 1 5 では、発話者がいかなる発話速度で話しても、発話内容に対応した動作スペクトルグラフを得ることができる。

【 0 0 6 1 】

また、口唇部分の上下左右端の 4 つの特徴点あるいは下顎を含めた 5 つの特徴点は、それぞれ独特の動きをするため、4 つあるいは 5 つの特徴点の動作履歴は、発音される母音によって大きく異なることになる。その結果、これらの特徴点の動作履歴グラフを変換した動作スペクトルグラフは、発音される母音によって大きく異なることになるので、発音される母音に応じて区別可能となる。例えば、イギリス英語には、母音が 2 4 音あると言われているが、このイギリス英語の母音（以下、「英母音」ともいう。）それぞれについても口唇動作の動作履歴をスペクトル解析することによって、動作スペクトルグラフに変換することができることが確認されている（尾上他：イギリス英母音に対する口唇動作解析、映像情報メディア学会年次大会、2009、7-3、尾上他：イギリス英母音に対する口唇動作解析（その 2）映像情報メディア学会冬期大会、2009、5-9）。

【 0 0 6 2 】

ここで、図7を参照して、同じ発話者が異なる母音列を示す複数の英単語を発話した際の動作スペクトルグラフの例について説明する。

図7では、英単語の中で、日本語の「ウ」に似た母音列を示す単語である「look」[luk]と「luke」[lu・k]、日本語の「ア」に似た母音列を示す単語である「father」[fa・ð・ər]、「pat」[pæt]および「cup」[cʌp]を同じ発話者が発話した際に、画像処理手段13で口唇左端部の特徴点を抽出し、動作測定手段14で動作履歴を測定したものをデータ変換手段15で変換して得られた動作スペクトルのグラフを例示している。

【0063】

図7に示すように、口唇左端部の動作スペクトルグラフでは、「ア」に似た母音列および「ウ」に似た母音列の双方で、周波数が“1”のところにピークが存在するが、「ウ」に似た母音列では強い動作スペクトルを示し、「ア」に似た母音列では弱い動作スペクトルを示している。このように、「ア」に似た母音列と「ウ」に似た母音列とでは、口唇左端の動作スペクトルグラフが大きく異なることになる。なお、ここでは図示しないが「ウ」に似た母音列の「look」と「luke」については、口唇上部における動作スペクトルグラフが異なる。このように、母音によって特徴的なスペクトルが得られる特徴点があり、それぞれ異なる動作スペクトルが得られる。

10

【0064】

次に、図8を参照して、日本語の「ア、イ、ウ、エ、オ」に近い各英母音について、それぞれ異なる単語を発話した際のそれぞれの動作スペクトルグラフの例について説明する。図8では、縦軸に周波数を取り、横軸には左側から順に、「ア」に近い英母音から「オ」に近い英母音を含む単語対を並べている。例えば、「ア」に近い英母音を含む単語対は、「father」と「cup」である。また、図8において、それぞれの単語対の下方に示す「口唇下部」、「口唇左端部」、「下顎部」は、その単語対で特徴的な動作スペクトルが表れる特徴点を示したものである。

20

【0065】

図8に示すように、例えば「father」の母音である[a:]と「cup」の母音である[ʌ]とでは、動作スペクトルの数値から、前者の口唇下部の動きが大きいことが分かる。同様に、「king」の母音である[i]と「seat」の母音である[i:]とでは、動作スペクトルの数値から、後者の口唇下部の動きが大きいことが分かる。

30

このように、似た母音列の単語間でも、異なる動作スペクトルが得られる。

以上のように、データ変換手段15によって、英母音のそれぞれについて英単語を発話した際の口唇動作の動作履歴をスペクトル解析することによって、英母音のそれぞれの特徴を表した動作スペクトルを得ることができる。データ変換手段15によって得られた動作スペクトルは、差分算出手段16に出力される。

【0066】

図2に戻って、差分算出手段16は、データ変換手段15で解析された、発話者がある単語を発話した際の動作スペクトルと、指導者が同じ単語を発話した際の模範的な動作スペクトルとの差分を算出するものである。

ここで、口唇動作は、動作スペクトルの複数のスペクトル成分のうち、主に第1成分に反映される。言い換えれば、動作スペクトルの第1成分は、口唇動作においては、特徴点の動きの大きさを示す情報となる。

40

【0067】

ここで、図9を参照して、複数の発話者が同じ母音（図9に示す例では、[ʌ]）を含む単語を発話した際にそれぞれ得られる口唇左端部の動作スペクトルグラフの例について説明する。

この図9（a）は、ネイティブが[ʌ]を母音として含む単語を発話した際の動作スペクトルグラフであり、（b）は、学習を始めたばかりの発話者が[ʌ]を母音として含む単語を発話した際の動作スペクトルグラフであり、（c）は、（b）と同じ発話者が、ある程度学習が進んだ時点で、[ʌ]を母音として含む単語を発話した際の動作スペクトルグラフである。

【0068】

図9（a）に示すように、ネイティブの動作スペクトルグラフは、動作スペクトル全体を100パーセントとしたときに、動作スペクトルの第1成分が動作スペクトル全体に占める比率が、おおよそ25パーセントとなっている。一方、図9（b）に示すように、学習を始めたばかりの発話者の動作スペクトルグラフは、動作スペクトル全体を100パーセントとしたときに、動作スペクトルの第1成分が動作スペクトル全体に占める比率が、おおよそ60パーセントとなっている。そして、図9（c）に示すように、ある程度学習が進んだ（b）と同じ発話者の動作スペクトルグラフは、動作スペクトルの複数の成分のうち、動作スペクトル全体を100パーセントとしたときに、動作スペクトルの第1成分が動作スペクトル全体に占める比率が、おおよそ40パーセントに減じている。

【0069】

図9（a）～（c）によれば、同じ母音を含む単語を発話した場合であっても、発話者間の口唇動作の違いによって、動作スペクトル全体に占める第1成分の比率に違いが生じることが分かる。また、学習が進み、口唇動作が改善されるにつれて、発話者の動作スペクトルの第1成分が動作スペクトル全体に占める比率が、ネイティブの動作スペクトルの第1成分が動作スペクトル全体に占める比率に近づいていることが分かる。

【0070】

以上のように、発話者（発話者）の口唇動作と指導者の口唇動作との差は、同じ母音を含む単語を発話した際に、それぞれの動作スペクトルの第1成分が動作スペクトル全体に占める比率の違いとなって表れる。

そして、発話者の前記比率と指導者の前記比率との差分が、発話者と指導者との発音の差を表すので、この差分を利用することにより、発話者の口唇動作の修正量を求めることが可能となる。

【0071】

そこで、差分算出手段16では、発話者が、発話内容指定手段12で指定された単語を発話した際の動作スペクトルの第1成分が動作スペクトル全体において占める比率から、指導者が同じ単語を発話した際の動作スペクトルの第1成分が動作スペクトル全体において占める比率を減算して得られる差分を特徴点ごとに算出する。

【0072】

なお、発話者の前記比率から指導者の前記比率を減算して算出された差分が負の値であるときは、指導者の模範的な口唇動作に対し、発話者の口の開き方が小さすぎることになる。同様に、算出された差分が正の値であるときは、指導者の模範的な口唇動作に対し、発話者の口の開き方が大きすぎることになる。このように、差分の値の正負によっても、口唇動作を修正する方向（大きい方向か小さい方向か）が分かる。

【0073】

以下では、特徴点ごとの差分を区別する場合、口唇上部の差分を U とし、口唇下部の差分を D とし、口唇左端部の差分を L とし、口唇右端部の差分を R とし、下顎の差分を J とし、さらに、画像処理手段13によって口唇中央部奥行き方向の特徴点が抽出されているときは、口唇中央部奥行き方向の差分を Depth と表すこととする。

差分算出手段16で算出された特徴点ごとの差分は、修正量算出手段17にそれぞれ出力される。

【0074】

10

20

30

40

50

なお、指導者の模範的な動作スペクトルは、予め指導者に、その言語の全ての母音（あるいは母音を含む単語であってもよい）を発話してもらい、そのときの指導者の口唇部分を撮影手段 20 でその都度撮影し、画像処理手段 13 によって、撮影手段 20 で取得された画像から予め設定した特徴点の位置を抽出し、さらに、動作測定手段 14 によって、特徴点の動作履歴を測定し、そして、データ変換手段 15 によって、動作履歴を数値解析することによって、母音ごとの動作スペクトルを求めることができる。

そして、差分算出手段 16 によって、動作スペクトルの第 1 成分が動作スペクトル全体において占める比率を、母音ごとに算出することができる。

【0075】

この指導者の前記比率は、ここでは、前記したように動作スペクトルと対応付けてデータ記憶手段 11 に記憶しているので、差分算出手段 16 は、発話者の動作スペクトルが入力された際に適宜読み出すこととする。

【0076】

修正量算出手段 17 は、差分算出手段 16 で算出された差分に基づいて、発話者の口唇動作の修正量を算出するものである。

修正量算出手段 17 は、差分算出手段 16 で算出された特徴点ごとの差分の絶対値を、特徴点ごとに予め定めた閾値と比較し、差分の絶対値が閾値よりも大きいと判定した特徴点については、特徴点ごとに予め定めた補正関数に従って、修正量を算出する。

一方、差分の絶対値が閾値よりも小さいと判断した特徴点については、指導者の特徴点の動作に十分近づいているものとして、発話者の特徴点の動作の修正を指示しない。この閾値は、予め実験等を行って母音ごとに一つの値を定めておく。

【0077】

ここで、図 10 を参照して、本実施形態で適用される補正関数について説明する。

図 10 は、口唇上部の差分 U に対する口唇上部の動作の修正量 f_1 との関係を示した補正関数のグラフを表しており、縦軸に口唇上部の動作の修正量 f_1 を取り、横軸に差分

U を取っている。図 10 において、 f_{1max} は、当該指定の母音の発音時に、口唇上部を最も大きく動かしたときの、発話者の口唇上部の動作の修正量を示すものであり、 f_{1min} は、当該指定の母音の発音時に、口唇上部を最も小さく動かしたときの、発話者の口唇上部の動作の修正量を示すものであり、 $f_1(U)$ は、この 2 点を直線的に結んだ補正式である。図 10 における f_{1max} は、最も大きく口唇上部を動かしたときの修正量であり、人によって異なる。 f_{1min} は、口唇上部を動かしていないときの修正量に相当する。本発明を実施する上で、これらの値を測定する必要は無いが、 f_{1max} 、 f_{1min} として (U) に上限、下限があることを概念的に示している。但し、 f_{1min} は、口唇左端部、口唇右端部、口唇上部、口唇下部、下顎部の各特徴点について、必ずしも動かしていない状態ではなく、反対方向に動いた場合も含む。例えば、口を横に広げるべき発音で、口をすばめた場合である。

【0078】

また、図 10 において、 $-Th_u$ は、口唇上部を大きく開ける方向に修正するかどうかを判定するための閾値であり、 $+Th_u$ は、口唇上部を小さく開ける方向に修正するかどうかを判定するための閾値である。差分 U の絶対値である $abs(U)$ が閾値 Th_u より小さい場合（閾値 $+Th_u \sim -Th_u$ 内に収まっているとき）、対応する修正量 $f_1(+Th_u) \sim f_1(-Th_u)$ を発話者に修正量として指示しないか、あるいは、 U を補正関数に代入する処理を行わないものとする。

【0079】

なお、人により口唇の物理的な大きさが異なり、口唇動作量も異なるため、前記した補正関数を発話者に適用する場合、予めデータ記憶手段 11 に記憶された指導者の口唇動作量を基準に正規化を行うこととする。その方法の一例として、口を自然に閉じているときの口唇左端部と口唇右端部との差、すなわち横幅と、口唇上部と口唇下部との差、すなわち縦幅とを、指導者および発明者のそれぞれについて求め、さらに、指導者と発明者の横幅同士と縦幅同士とをそれぞれ比較して、それぞれ比率を求めておく。そして、このよう

10

20

30

40

50

にしてそれぞれ求めた比率を用いて、発話学習時に測定された発話者の口唇動作量を正規化する。

【0080】

そして、発話者に正規化された補正関数 $f_1(U)$ に従い、図10に示すように直線的に補正する場合、差分 U が予め定めた閾値 $+Th_u$ よりも大きいときは、閾値 $+Th_u$ から離れるにつれて、発話者の口唇上部の動きを小さくする方向に、口唇上部の動きの修正量を大きくする。一方、 U が閾値 $-Th_u$ よりも小さいときは、閾値 $-Th_u$ から離れるにつれて口唇上部の動きを大きくする方向に、口唇上部の動きの修正量を大きくする。このようにして、補正関数 $f_1(U)$ に従い、発話者の口唇上部の動作を指導者の口唇上部の動作に近づけるための、発話者の口唇上部の動作の修正の方向と大きさを特定した修正量を算出することができる。

10

【0081】

また例えば、口唇左端部の動作の修正量を求める場合、発話者に正規化された口唇差端部の補正関数に従い、図10に示すように直線的に補正する場合、差分 L (図示せず) が予め定めた閾値 Th_l (図示せず) よりも大きいときは、閾値 Th_l (図示せず) から離れるにつれて、発話者の口唇左端部の動きを小さくする方向に、口唇左端部の動きの修正量を大きくする。一方、 L (図示せず) が予め定めた閾値 $-Th_l$ (図示せず) よりも小さいときは、閾値 $-Th_l$ (図示せず) から離れるにつれて口唇左端部の動きを大きくする方向に、口唇左端部の動きの修正量を大きくする。

【0082】

20

さらに例えば、口唇中央部の奥行き方向の動作の修正量を求める場合、発話者に正規化された口唇中央部の奥行き方向の補正関数に従い、図10に示すように直線的に補正する場合、差分 $Depth$ (図示せず) が予め定めた閾値 Th_{depth} (図示せず) よりも大きいときは、閾値 Th_{depth} (図示せず) から離れるにつれて、発話者の口唇中央部をすぼめる方向に、口唇中央部の奥行き方向の動きの修正量を大きくする。一方、 $Depth$ (図示せず) が予め定めた閾値 $-Th_{depth}$ (図示せず) よりも小さいときは、閾値 $-Th_{depth}$ (図示せず) から離れるにつれて口唇中央部を突き出す方向に、口唇中央部の奥行き方向の動きの修正量を大きくする。他の特徴点についても、同様に、それぞれ対応する補正関数に従い、修正量を算出することができる。修正量算出手段17は、算出した修正量を、修正情報出力手段18に出力する。

30

【0083】

修正情報出力手段18は、修正量算出手段17によって算出された修正量を、発話者が認識可能な形式の修正情報として出力する手段である。この修正情報は、特徴点の修正の方向と大きさを示すものである。ここでは、予め修正量に応じて生成されたパターンを図示しない記憶手段に記憶しておき、修正情報出力手段18によって適宜読み出すこととする。ただし、修正情報出力手段18は、修正量算出手段17によって算出された修正量に応じて、パターンをその都度生成する機能を有していてもよい。

【0084】

修正の向きは、口唇上部を、さらに上へ動かすように修正する場合は、上方向の矢印で、上への動きを小さくするように修正するときは、下方向の矢印で示す。以下同様に、口唇下部を、さらに下へ動かすように修正する場合は、下方向の矢印で、下への動きを小さくするように修正するときは、上方向の矢印で示す。口唇左端部を、口をすぼめる方向に動かすように修正するときは右方向の矢印で、開く方向に動かすように修正するときは左方向の矢印で示す。逆に、口唇右端部を、口をすぼめるように動かすように修正するときは左方向の矢印で、開く方向に動かすように修正するときは右方向の矢印で示す。下顎部を、さらに下へ動かすように修正するときは、下方向の矢印で、下への動きを小さくするように修正するときは、上方向の矢印で示す。口唇中央部を、より突き出すように修正するときは、平面的に表示する場合、突き出すような矢印を透視図的に描いて示し、立体表示の場合、視差をつけて飛び出す矢印で示す。口唇中央部を、あまり突き出さないように修正するときは、逆方向の矢印で示す。この矢印の大きさは、補正量によって異なり、補

40

50

正量の大きさにより大中小、大小など経験的に決めてよい。

【 0 0 8 5 】

このパターンは、発話者が把握可能な形式であり、かつ、特徴点の修正の方向と大きさを示すことができればどのようなものであってもよい。例えば、パターンを、表示装置 30 のスピーカ（図示せず）で再生可能な音声としてもよいし、表示装置 30 の表示画面 31 に表示可能な態様のテキストとしてもよい。

【 0 0 8 6 】

また例えば、パターンを修正の方向と大きさを示す図形（例えば矢印）の C G（Computer Graphics）とし、この図形の C G を、発話者の口唇部分の画像における、修正量算出手段 17 によって修正量が算出された特徴点に対応する位置に合成した合成画像を、修正情報としてもよい。この図形を矢印とする場合、矢印の向きで修正の方向を示し、矢印の長さで修正の大きさを示すことができる。例えば、矢印が長くなるほど修正量が大きく、短くなるほど修正量が小さいことを示す。矢印の長さで修正量を表す場合、予め修正量に応じて矢印の長さを設定しておく。

そして、この矢印の C G を発話者の口唇部分の画像に合成する場合、画像における動作を修正する特徴点に対応する位置に矢印の根元の位置を合わせるとよい。このようにすると、発話者に、矢印の向き（矢印が口唇の内側を向いているか外側を向いているか）によって、口唇のどの部分をどのように修正すればよいかを直感的に理解させやすくなる。

【 0 0 8 7 】

なお、表示画面 31 が、画像を立体的に表示可能な場合には、パターンを立体的な図形の C G としてもよい。これによれば、修正量算出手段 17 により、口唇中央部の奥行き方向の動作の修正量が算出された場合に、修正情報出力手段 18 によって、口唇中央部に、修正の方向と大きさを示す立体的な矢印の C G を合成して修正情報とすることができる。この場合、矢印の先端を画面手前側に向けるか画面奥側に向けるかによって、修正の方向を示すことができる。

【 0 0 8 8 】

このように、修正情報を合成画像とする場合、修正情報出力手段 18 は、修正量算出手段 17 から、ある特徴点の修正量の入力を受け付けると、撮影手段 20 から発話者の口唇部分の画像を取得すると共に、図示しない記憶手段からその特徴点の修正量に応じた図形の C G を読み出して、発話者の口唇部分の画像に当該図形の C G を合成することによって合成画像を生成する。また、修正情報を予め図示しない記憶手段に記憶したパターンとする場合、修正情報出力手段 18 は、修正量算出手段 17 から、ある特徴点の修正量の入力を受け付けると、図示しない記憶手段からその特徴点の修正量に応じたパターン（音声、テキスト等）を読み出す。

このようにして、修正情報出力手段 18 によって生成され、あるいは、図示しない記憶手段から読み出された修正情報は、表示装置 30 に出力される。

【 0 0 8 9 】

そして、修正情報出力手段 18 から表示装置 30 に出力した修正情報を、表示装置 30 の表示画面 31 に表示し、あるいは、表示装置 30 のスピーカ（図示せず）によって再生することで、発話者に、口唇動作の改善点を客観的に認識させることが可能となる。

【 0 0 9 0 】

ここで、図 11 を参照して、修正情報出力手段 18 から修正情報を表示装置 30 に出力し、この修正情報を表示画面 31 に表示させた際の画面構成について説明する。ここでは、発話学習支援装置 1 は、修正量算出手段 17 によって、発話者の口唇左端部の口唇動作が指導者の口唇左端部の動作よりも大きすぎると判断し、発話者の口唇左端部の動作を小さく動作させるための修正量を算出し、修正情報出力手段 18 によって、この修正量に応じた修正情報を表示装置 30 に出力したものとする。

【 0 0 9 1 】

ここでは、発話学習支援装置 1 は、修正情報出力手段 18 によって、発話者の口唇部分の画像上における口唇左端部に対応する位置に、修正の大きさと方向を示す図形（ここで

10

20

30

40

50

は、矢印)のCGを合成した画像と、修正の大きさと方向を示すテキストと、を修正情報として、表示装置30に出力したものとする。

【0092】

図11に示すように、表示画面31には、発話学習支援装置1の修正情報出力手段18から出力された修正情報として、発話者の口唇部分の画像における口唇左端部に対応する位置に、口唇左端部から口唇右端部方向に向かって伸びる所定長さの矢印Yが合成された画像Gsが表示されている。またさらに、画像Gsの下方には、発話学習支援装置1の修正情報出力手段18から出力された修正情報として、「口唇の左端をもう少し小さく動かしてください」というテキストデータTbが表示されている。このようにして、発話学習支援装置1の修正情報出力手段18から出力された修正情報を表示装置30の表示画面31に表示させることで、発話者に、自身の口唇動作の改善点を客観的に認識させることが可能となる。なお、ある特徴点の動きを修正した結果、他の特徴点の動きが指導者の動きから外れることもあり得る。その場合、その特徴点について差分が閾値内に収まっているかどうか判定し、収まっていない場合、修正量を演算して修正情報として発話者に示し、修正情報を確認した発話者が再度発話したときに、差分が閾値内に収まっているかどうか再度判定することとし、最終的に全ての特徴点の動きと指導者の対応する特徴点の動きとの差分が閾値内に収まるように、発話者はトレーニングを続ける。

10

【0093】

以上に説明した発話学習支援装置1は、コンピュータにおいて各手段を各機能プログラムとして実現することも可能であり、各機能プログラムを結合して、発話学習支援プログラムとして動作させることも可能である。

20

【0094】

[発話学習支援装置の動作]

次に、図12を参照して、発話学習支援装置1を含む発話学習支援システムの動作について説明する。

発話学習支援システムは、発話学習支援装置1の発話内容指定手段12によって、表示装置30から、発話者により、ある言語の学習開始の決定がされたことを示す信号の入力を受け付ける(ステップS11)。そして、発話学習支援システムは、発話学習支援装置1の発話内容指定手段12によって、ステップS11において入力を受け付けた信号に応じて、データ記憶手段11に記憶されたその言語の複数のデータ組に含まれる単語の中から、発話者に発話させる単語を一つ選択し、その単語の発話を指示するテキストデータと、指導者がその単語を発話した際の口唇部分の画像と、を読み出し、表示装置30の表示画面31に表示する(ステップS12)。なおこのとき、データ記憶手段11に、その単語の音声データがさらに記憶されている場合、発話内容指定手段12は、当該音声データをさらに読み出して表示装置30のスピーカ(図示せず)から再生してもよい。

30

【0095】

続いて、発話学習支援システムは、撮影手段20によって、発話者が、ステップS12において指示された単語を発話した際の口唇部分を撮影する(ステップS13)。

そして、発話学習支援システムは、発話学習支援装置1の画像処理手段13によって、ステップS13において撮影された発話者の口唇部分の画像の入力を受け付けると、当該口唇部分の画像から口唇動作の基準となる予め設定した特徴点の位置を抽出する(ステップS14)。

40

【0096】

そして、発話学習支援システムは、発話学習支援装置1の動作測定手段14によって、ステップS14において位置が抽出された特徴点の動作履歴を測定する(ステップS15)。そして、発話学習支援装置1は、データ変換手段15によって、ステップS15において測定された特徴点ごとの動作履歴をフーリエ解析し、特徴点ごとの動作スペクトルを生成する(ステップS16)。

【0097】

さらに、発話学習支援システムは、発話学習支援装置1の差分算出手段16によって、

50

ステップS16において生成された特徴点ごとの動作スペクトルグラフにおける動作スペクトルの第1成分が動作スペクトル全体に占める比率を求め、この比率と、予めデータ記憶手段11に記憶された指導者の模範的な動作スペクトルグラフから求めた動作スペクトルの第1成分が動作スペクトル全体に占める比率と、の差分を特徴点ごとに算出する(ステップS17)。

【0098】

そして、発話学習支援システムは、発話学習支援装置1の修正量算出手段17によって、ステップS17において算出された特徴点ごとの差分の絶対値が、発話者が発話した単語の母音について予め定めた閾値より大きいかを、特徴点ごとに判定する(ステップS18)。そして、予め定めた閾値より差分の絶対値の方が大きい特徴点があると判定された場合(ステップS18でYes)、発話学習支援装置1は、修正量算出手段17によって、当該特徴点の差分を、予め定めた補正関数に代入し、差分に応じた、修正の方向と大きさを特定した修正量を算出して(ステップS19)、そのままステップS20に進む。

一方、予め定めた閾値より差分の絶対値の方が大きい特徴点がないと判定された場合(ステップS18でNo)、発話学習支援装置1は、修正量算出手段17によって修正量を算出せずに、そのまま処理を終了する。

【0099】

そして、発話学習支援システムは、発話学習支援装置1の修正情報出力手段18によって、ステップS19において修正量が算出された特徴点について、当該修正量に応じた修正情報を図示しない記憶手段から読み出し、あるいは、当該修正量に応じた修正情報を生成して、表示装置30に出力する(ステップS20)。

【0100】

そして、発話学習支援システムは、表示装置30によって、ステップS20において発話学習支援装置1から出力された修正情報の入力を受け付ける。そして、発話学習支援システムは、表示装置30によって入力を受け付けた修正情報が、画像またはテキストの場合、当該修正情報を表示画面31に表示し、音声の場合、スピーカ(図示せず)から再生する(ステップS21)。

そして、発話学習支援システムは、ステップS13に戻り、例えば表示画面31に表示された修正情報を参照して、発話内容指定手段12によってすでに指定されている単語を発話者が繰り返し発話した際の口唇部分を撮影手段20によって撮影する。このようにして、発話学習支援システムは、ステップS18において予め定めた閾値より差分の絶対値の方が大きい特徴点がないと判定されるまで、ステップS13からステップS21を繰り返す。

以上のようにして、発話学習支援システムは、発話者に口唇動作の修正量を、発話者が客観的に認識可能な態様で示すことができる。

【0101】

本実施形態に係る発話学習支援装置1によれば、発話者の口唇動作を指導者の口唇動作と比較した結果として、発話者の口唇動作の修正量を、発話者が客観的に認識可能な態様で示すことができる。このため、発話者に、自身の口唇動作の改善点を客観的に認識させることが可能となる。これにより、効果的な学習教育に役立てることができる。

また、本実施形態に係る発話学習支援装置1によれば、聾話者等、生まれつき難聴もしくは耳が聞こえない人が、口話を学習する際に正しい発音を身に着けるのに役立てることができる。

さらに、本実施形態に係る発話学習支援装置1は、脳梗塞や脳卒中などによる片麻痺により、口唇の左部、あるいは、右部の動作に支障をきたすようになった患者が、単語を正しく発音するための口唇動作のリハビリに利用することができる。

【0102】

以上、本発明の実施形態について説明したが、本発明は前記実施形態に限定されるものではなく、本発明の趣旨を逸脱しない範囲で適宜変更可能である。

例えば、前記実施形態では、修正量算出手段17で修正量を求める際に適用する補正関

10

20

30

40

50

数を図 10 に示すような一次式としたが、これに限られず、差分の絶対値が閾値から大きく外れたときの修正量を大きくし、差分の絶対値が閾値に近いときの修正量を小さくする場合には、2 次の多項式を用いればよい。また、差分の絶対値が閾値に近いときの微妙な口唇動作の修正を優先させる場合には、平方根（0.5 次式）や、対数の多項式を用いればよい。

【0103】

また、前記実施形態では、データ記憶手段 11 に、単語をテキスト形式で記憶しておき、発話内容指定手段 12 によってデータ記憶手段 11 から単語のテキストデータを読み出して表示装置 30 の表示画面 31 に表示することにより、発話者に発話させる単語を指示していたが、これに限られず、データ記憶手段 11 に、単語に対応付けて、当該単語を音声合成した音声データを記憶しておき、発話内容指定手段 12 によってこの音声データを読み出して表示装置 30 に内蔵されたスピーカ（図示せず）から再生することにより、発話者に発話させる単語を指示してもよい。また、音声とテキストの両方を用いてもよい。

10

【0104】

なお、撮影手段 20 としてステレオカメラを用いた場合、例えば、ステレオカメラを構成する左右カメラの口唇下部の水平方向の座標差を視差として計算し、予め分かっているステレオカメラのカメラ間距離等のカメラパラメータを適用して、ステレオ測定の原理により口唇下部の奥行き方向の動きを算出する。口唇上端の奥行き方向の動きも同様にして算出することができる。

【0105】

20

次に、図 13 を参照して、前記実施形態の変形例に係る発話学習支援装置 1B について説明する。変形例に係る発話学習支援装置 1B は、前記実施形態に係る発話学習支援装置 1 の構成に加え、評価値算出手段 19 を備えている。以下の説明では、前記実施形態と重複する構成要素については、同一の符号を付して説明を省略する。

【0106】

図 13 に示すように、発話学習支援装置 1B は、データ記憶手段 11 と、発話内容指定手段 12 と、画像処理手段 13 と、動作測定手段 14 と、データ変換手段 15 と、差分算出手段 16 と、修正量算出手段 17 と、修正情報出力手段 18 と、評価値算出手段 19 とを備えている。

【0107】

30

評価値算出手段 19 は、発話者の口唇動作を、指導者の口唇動作と比較した結果に基づいて、発話者の口唇動作の評価値を算出するものである。

評価値算出手段 19 は、ここでは、式（1）に示すように、データ変換手段 15 で得られた発話者の動作スペクトルグラフを構成する全スペクトルにそれぞれの重みを乗算して総スペクトルを計算し、さらに、データ記憶手段 11 に記憶された指導者の動作スペクトルグラフを構成する全スペクトルにそれぞれの重みを乗算して総スペクトルを計算し、指導者に対して発話者に動作スペクトルの第 1 成分が総スペクトルに対して占める量を評価値 X として算出する。

【0108】

【数 1】

40

$$X = \sqrt{\left(\frac{T_1}{\sum_{i=1}^N W_i T_i} - \frac{P_1}{\sum_{i=1}^N W_i P_i} \right)^2} \quad \dots (1)$$

【0109】

式（1）において、 T_1 は、指導者の動作スペクトルの第 1 成分を示し、 P_1 は、発話者（学習者）の動作スペクトルの第 1 成分を示す。また、式（1）において、重み W_i は

50

、適宜設定することができるが、動作スペクトルの第1成分に適用される重み W_1 は、他の成分の重みよりも大きくすることとする。

【0110】

このように評価値算出手段19によって、評価値 X を算出することで、練習前後での評価値の変化によって、発話者に、自身の口唇動作が指導者の口唇動作に近づいているかどうか客観的に認識させることが可能となる。

評価値算出手段19によって算出された評価値 X は、表示装置30に出力されて表示画面31に表示される。この評価値 X を、修正情報出力手段18から表示装置30に出力された修正情報と合わせて表示画面31に表示してもよいし、いずれか一方のみを表示画面31に表示してもよい。この評価値 X を修正情報と合わせて表示画面31に表示すると、発話者に、自身の口唇動作の改善点をより認識させやすくなるため好ましい。

10

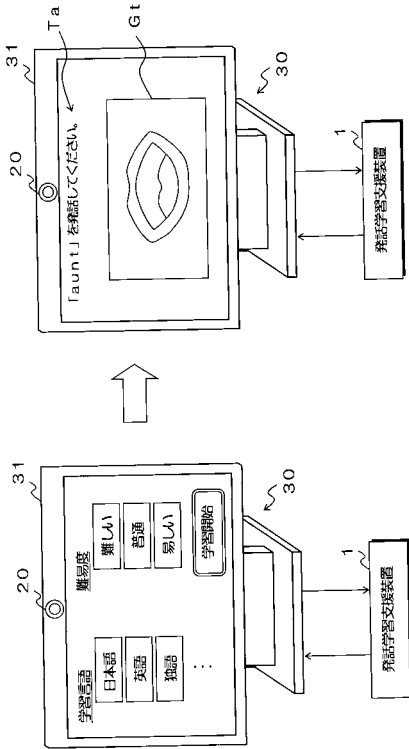
【符号の説明】

【0111】

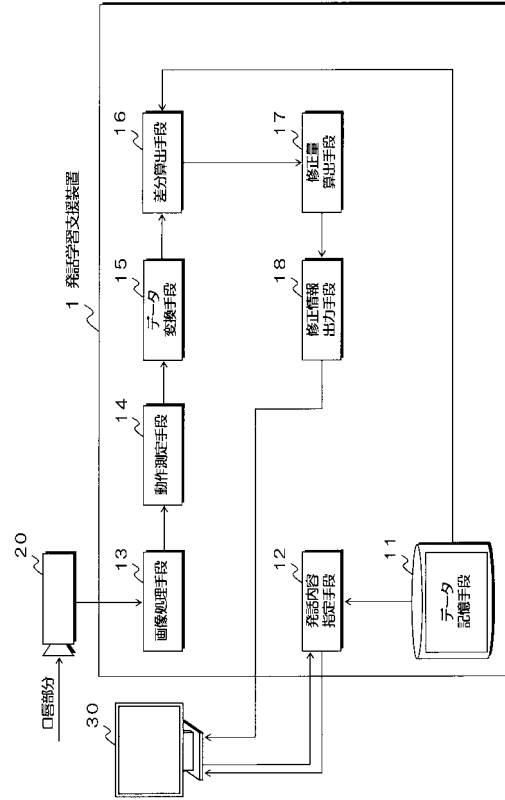
- | | |
|------|----------|
| 1、1B | 発話学習支援装置 |
| 11 | データ記憶手段 |
| 12 | 発話内容指定手段 |
| 13 | 画像処理手段 |
| 14 | 動作測定手段 |
| 15 | データ変換手段 |
| 16 | 差分算出手段 |
| 17 | 修正量算出手段 |
| 18 | 修正情報出力手段 |
| 19 | 評価値算出手段 |
| 20 | 撮影手段 |
| 30 | 表示装置 |
| 31 | 表示画面 |

20

【図1】



【図2】



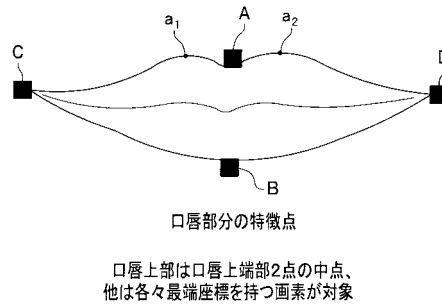
【図3】

データ認識手段のデータ構造例

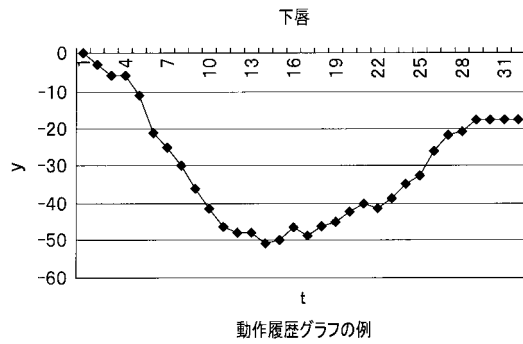
母音	単語	テキスト	画像	スペクトル	比率	難易度
a	aunt	テキスト1	画像1	スペクトル1	比率1	普通
	cup	テキスト2	画像2	スペクトル2	比率2	難しい
a:	bar	テキスト3	画像3	スペクトル3	比率3	易しい
	bark	テキスト4	画像4	スペクトル4	比率4	普通
i:	seat	テキスト5	画像5	スペクトル5	比率5	普通

...

【図4】

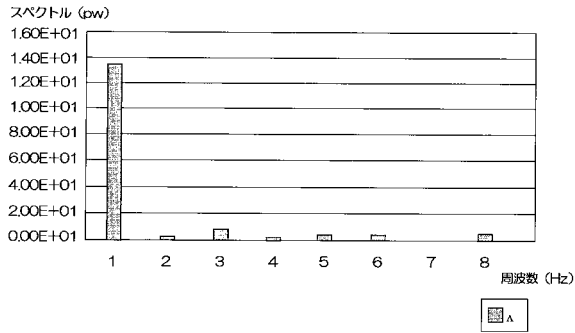


【図5】

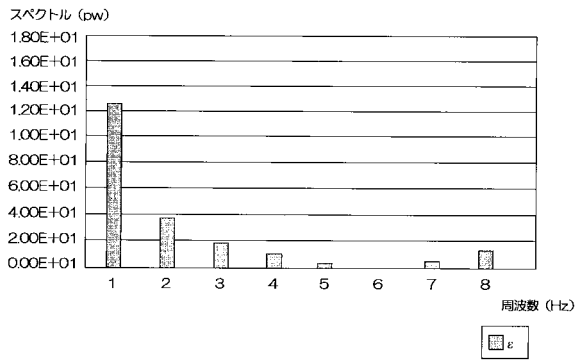


【図6】

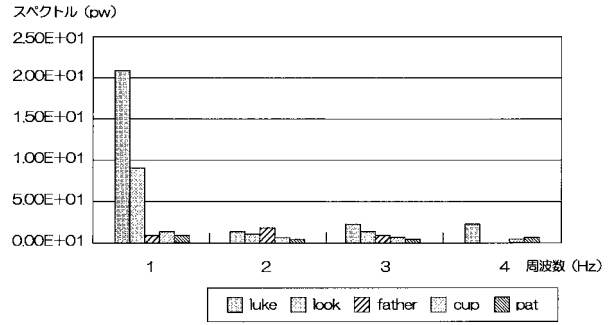
(a)



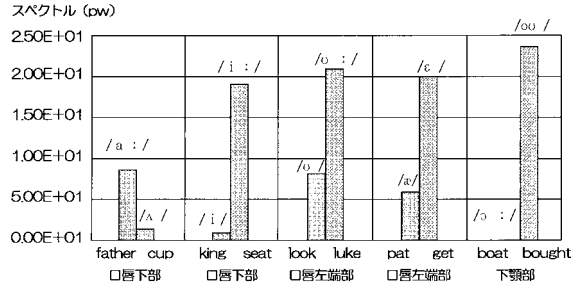
(b)



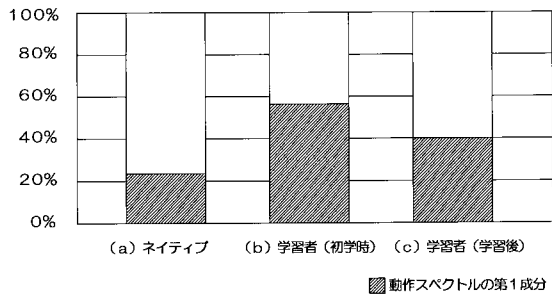
【図7】



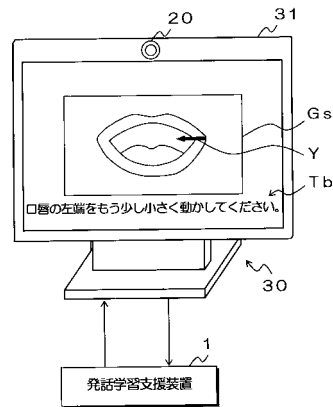
【図8】



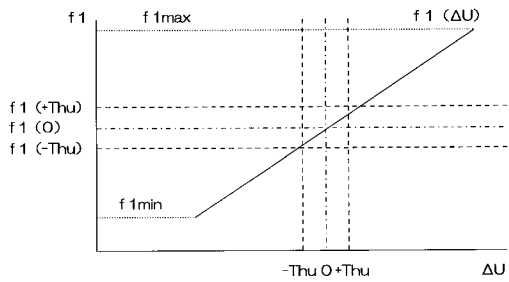
【図9】



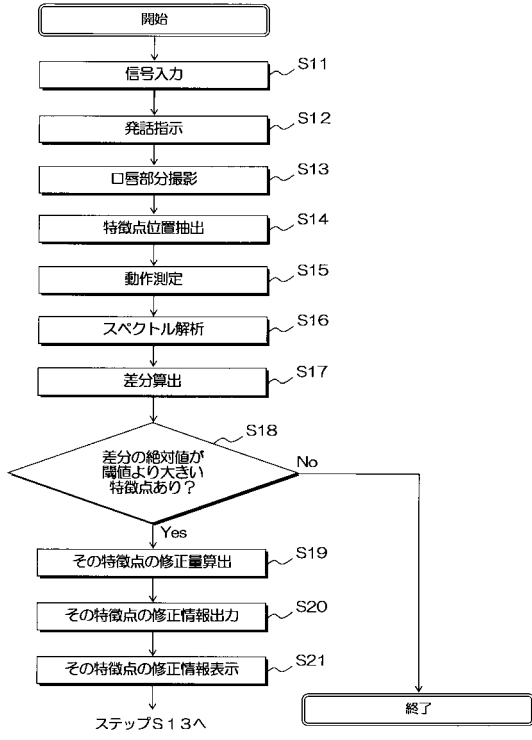
【図11】



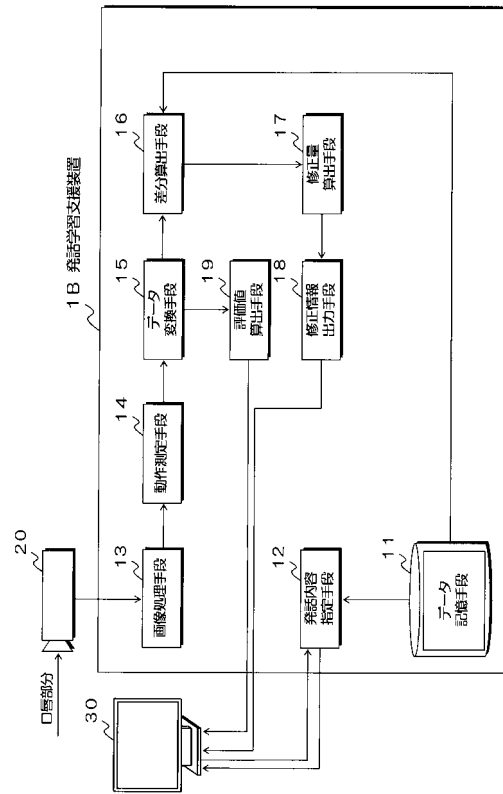
【図10】



【図12】



【図13】



フロントページの続き

(72)発明者 齋藤 翼
神奈川県平塚市北金目四丁目1番1号 東海大学内

審査官 櫻井 茂樹

(56)参考文献 特開2008-158055(JP,A)
特開2006-162760(JP,A)
特開平03-098078(JP,A)
小田 まり子(外2名),見真似発音練習システムにおける口唇動作の誤り指摘,情報処理学会研究報告,日本,社団法人情報処理学会,2009年 2月20日,Vol.2009,No.15,P.185-192
尾上 拓(外2名),発声時の口唇動作スペクトルによるイギリス英語母音認識法の提案,電子情報通信学会2010年総合大会講演論文集 基礎・境界,社団法人電子情報通信学会,2010年 3月 2日,P.203
齋藤 翼(外2名),口唇動作による母音列認識の統合的な手法の検討,電子情報通信学会2010年総合大会講演論文集 基礎・境界,社団法人電子情報通信学会,2010年 3月 2日,P.204

(58)調査した分野(Int.Cl.,DB名)
G09B1/00~9/56、17/00~19/26