

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2019-197236
(P2019-197236A)

(43) 公開日 令和1年11月14日(2019.11.14)

(51) Int.Cl.		F I		テーマコード (参考)
G09B 19/04 (2006.01)		G09B	19/04	2C028
G09B 5/02 (2006.01)		G09B	5/02	5L096
G09B 19/00 (2006.01)		G09B	19/00	H
G06T 7/20 (2017.01)		G06T	7/20	300B
G09B 21/00 (2006.01)		G09B	21/00	G
審査請求 未請求 請求項の数 8 O L 公開請求 (全 14 頁)				

(21) 出願番号 特願2019-148071 (P2019-148071)
(22) 出願日 令和1年8月9日 (2019.8.9)

(71) 出願人 397022911
学校法人甲南学園
兵庫県神戸市東灘区岡本8丁目9番1号
(74) 代理人 100124039
弁理士 立花 顕治
(74) 代理人 100179213
弁理士 山下 未知子
(74) 代理人 100170542
弁理士 榎田 剛
(74) 代理人 100150072
弁理士 藤原 賢司
(72) 発明者 北村 達也
兵庫県神戸市東灘区岡本8丁目9番1号
甲南大学内
Fターム(参考) 2C028 BA03 BB04 BC02
5L096 BA08 CA04 DA03 DA04 HA04

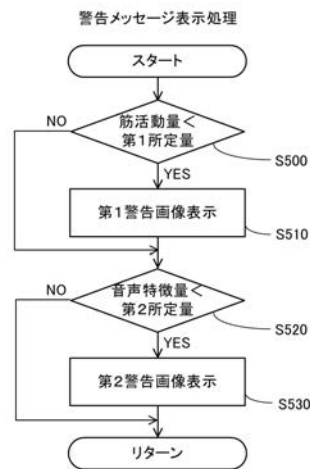
(54) 【発明の名称】 発話訓練システム、発話訓練方法及びプログラム

(57) 【要約】

【課題】より効果的に被訓練者の発話訓練を行なうことが可能な発話訓練システム、発話訓練方法及びプログラムを提供する。

【解決手段】発話訓練システムは、被訓練者の発話訓練に用いられる。発話訓練システムは、撮像手段と、表示手段とを備える。撮像手段は、被訓練者の顔を撮像し、動画像データを生成する。表示手段は、動画像データが示す動画像を表示する。表示手段は、被訓練者の口の動きの量を示す画像を動画像に重畳して表示する。この発話訓練システムによれば、被訓練者が口の周りの筋肉を大きく動かすことを意識して発話訓練を行なうことができるため、より効果的に被訓練者の発話訓練を行なうことができる。

【選択図】 図9



【特許請求の範囲】**【請求項 1】**

被訓練者の発話訓練に用いられる発話訓練システムであって、
前記被訓練者の顔を撮像し、動画像データを生成する撮像手段と、
前記動画像データが示す動画像を表示する表示手段とを備え、
前記表示手段は、前記被訓練者の口の動きの量を示す画像を前記動画像に重畳して表示する、発話訓練システム。

【請求項 2】

前記表示手段は、前記被訓練者に音読させる文章をさらに表示する、請求項 1 に記載の発話訓練システム。

【請求項 3】

前記表示手段は、前記被訓練者の発話に関する評価結果をさらに表示する、請求項 1 又は請求項 2 に記載の発話訓練システム。

【請求項 4】

前記表示手段は、前記被訓練者の発話が所定要件を満たさない場合に、警告メッセージをさらに表示する、請求項 1 から請求項 3 のいずれか 1 項に記載の発話訓練システム。

【請求項 5】

前記口の動きの量を示す画像は、前記口が移動した軌跡を示す線である、請求項 1 から請求項 4 のいずれか 1 項に記載の発話訓練システム。

【請求項 6】

前記動画像データに基づいてオプティカルフローを算出する算出手段と、
前記オプティカルフローに基づいて前記口の動きの量を示す画像を生成する生成手段とをさらに備える、請求項 1 から請求項 5 のいずれか 1 項に記載の発話訓練システム。

【請求項 7】

発話に関して被訓練者を訓練する発話訓練方法であって、
前記被訓練者の顔を撮像し、動画像データを生成するステップと、
前記動画像データが示す動画像を表示するステップと、
前記被訓練者の口の動きの量を示す画像を前記動画像に重畳して表示するステップとを含む、発話訓練方法。

【請求項 8】

被訓練者の発話訓練に用いられるプログラムであって、
撮像手段に、前記被訓練者の顔を撮像させ、動画像データを生成させるステップと、
表示手段に、前記動画像データが示す動画像を表示させるステップと、
前記表示手段に、前記被訓練者の口の動きの量を示す画像を前記動画像に重畳して表示させるステップとをコンピュータに実行させるプログラム。

【発明の詳細な説明】**【技術分野】****【0001】**

本発明は、発話訓練システム、発話訓練方法及びプログラムに関する。

【背景技術】**【0002】**

特開平 7 - 3 1 9 3 8 0 号公報（特許文献 1）は、発声訓練装置を開示する。この発声訓練装置においては、被訓練者の発声の調音法とモデル発声の調音法とのずれに基づく指示文が被訓練者にフィードバックされる。この発声訓練装置によれば、被訓練者は、指示文に従って訓練を進めることで調音法の矯正を効果的に行なうことができる（特許文献 1 参照）。

【先行技術文献】**【特許文献】****【0003】**

【特許文献 1】 特開平 7 - 3 1 9 3 8 0 号公報

10

20

30

40

50

【発明の概要】**【発明が解決しようとする課題】****【0004】**

上記特許文献1に開示されている発声訓練装置においては、被訓練者が発声した音声に基づいて被訓練者へのフィードバックが行なわれている。しかしながら、被訓練者が発声した音声に基づいたフィードバックのみでは、必ずしも被訓練者の発話訓練が効果的に行なわれないことを本発明者は見出した。

【0005】

本発明は、このような問題を解決するためになされたものであって、その目的は、より効果的に被訓練者の発話訓練を行なうことが可能な発話訓練システム、発話訓練方法及びプログラムを提供することである。

10

【課題を解決するための手段】**【0006】**

本発明のある局面に従う発話訓練システムは、被訓練者の発話訓練に用いられる。発話訓練システムは、撮像手段と、表示手段とを備える。撮像手段は、被訓練者の顔を撮像し、動画像データを生成する。表示手段は、動画像データが示す動画像を表示する。表示手段は、被訓練者の口の動きの量を示す画像を動画像に重畳して表示する。

【0007】

本発明者は、口の周りの筋肉を大きく動かすことを意識して発話訓練を行なうと、音声器官の可動域が広がり、被訓練者が発する音声の明瞭性が向上することを見出した。この発話訓練システムによれば、被訓練者の口の動きの量を示す画像が動画像に重畳して表示されるため、口の動きが不十分か否かを被訓練者に視覚的に認識させることができる。その結果、この発話訓練システムによれば、被訓練者が口の周りの筋肉を大きく動かすことを意識して発話訓練を行なうことができるため、より効果的に被訓練者の発話訓練を行なうことができる。

20

【0008】

上記発話訓練システムにおいて、表示手段は、被訓練者に音読させる文章をさらに表示してもよい。

【0009】

この発話訓練システムによれば、被訓練者に音読させる文章が表示されるため、被訓練者は、表示される文章を音読するだけで発話訓練を行なうことができる。

30

【0010】

上記発話訓練システムにおいて、表示手段は、被訓練者の発話に関する評価結果をさらに表示してもよい。

【0011】

この発話訓練システムによれば、被訓練者の発話に関する評価結果が表示されるため、被訓練者は、評価結果を確認しながら発話訓練を行なうことができる。

【0012】

上記発話訓練システムにおいて、表示手段は、被訓練者の発話が所定要件を満たさない場合に、警告メッセージをさらに表示してもよい。

40

【0013】

この発話訓練システムによれば、被訓練者の発話が所定要件を満たさない場合に警告メッセージが表示されるため、被訓練者は、自らの発話が所定要件を満たしていないことを視覚的に認識することができる。

【0014】

上記発話訓練システムにおいて、口の動きの量を示す画像は、口が移動した軌跡を示す線であってもよい。

【0015】

上記発話訓練システムは、動画像データに基づいてオブティカルフローを算出する算出手段と、オブティカルフローに基づいて口の動きの量を示す画像を生成する生成手段とを

50

さらに備えてもよい。

【0016】

本発明の別の局面に従う発話訓練方法は、発話に関して被訓練者を訓練する。発話訓練方法は、被訓練者の顔を撮像し、動画像データを生成するステップと、動画像データが示す動画像を表示するステップと、被訓練者の口の動きの量を示す画像を動画像に重畳して表示するステップとを含む。

【0017】

この発話訓練方法によれば、被訓練者の口の動きの量を示す画像が動画像に重畳して表示されるため、口の動きが不十分か否かを被訓練者に視覚的に認識させることができる。その結果、この発話訓練方法によれば、より効果的に被訓練者の発話訓練を行なうことができる。

10

【0018】

本発明の別の局面に従うプログラムは、被訓練者の発話訓練に用いられる。プログラムは、撮像手段に、被訓練者の顔を撮像させ、動画像データを生成させるステップと、表示手段に、動画像データが示す動画像を表示させるステップと、表示手段に、被訓練者の口の動きの量を示す画像を動画像に重畳して表示させるステップとをコンピュータに実行させる。

【0019】

このプログラムがコンピュータによって実行されると、被訓練者の口の動きの量を示す画像が動画像に重畳して表示されるため、口の動きが不十分か否かを被訓練者に視覚的に認識させることができる。その結果、このプログラムによれば、より効果的に被訓練者の発話訓練を行なうことができる。

20

【発明の効果】

【0020】

本発明によれば、より効果的に被訓練者の発話訓練を行なうことが可能な発話訓練システム、発話訓練方法及びプログラムを提供することができる。

【図面の簡単な説明】

【0021】

【図1】スマートフォンを用いた発話訓練風景の一例を示す図である。

【図2】スマートフォンのハードウェア構成の一例を示す図である。

30

【図3】制御部によって実現される各ソフトウェアモジュールの関係の一例を示す図である。

【図4】動画表示処理の実行手順を示すフローチャートである。

【図5】ディスプレイに表示される画像の一例を示す図である。

【図6】オプティカルフロー表示処理の実行手順を示すフローチャートである。

【図7】筋活動量表示処理の実行手順を示すフローチャートである。

【図8】音声特徴量表示処理の実行手順を示すフローチャートである。

【図9】警告メッセージ表示処理の実行手順を示すフローチャートである。

【図10】ディスプレイに表示される画像の一例を示す図である。

【図11】訓練前後に録音した音声の振幅を示す図である。

40

【図12】訓練前後に録音した音声の基本周波数の変化幅を示す図である。

【図13】訓練前後に計測したV A Sを示す図である。

【発明を実施するための形態】

【0022】

以下、本発明の実施の形態について、図面を参照しながら詳細に説明する。なお、図中同一又は相当部分には同一符号を付してその説明は繰り返さない。

【0023】

[1. 概要]

本発明者の調査によって、健常者である大学生及び大学院生の約3割が発話のしにくさを自覚していることが分かった。本発明者が種々の発話訓練方法を試したところ、被訓練

50

者が細い棒を咥えた状態で発話訓練を行なうことによって、高い訓練効果が得られる可能性があることが分かった。特に、本発明者は、発話訓練時に、被訓練者が大きい声を出すとともに顔面の筋肉をしっかりと動かすことによって高い訓練効果が得られることを見出した。

【0024】

図1は、本実施の形態に従うスマートフォン100を用いた発話訓練風景の一例を示す図である。図1に示されるように、発話訓練において、被訓練者10は、棒20を咥えた状態で発声する。被訓練者10は、スマートフォン100に表示される画像を見ながら発話訓練を行なう。詳細については後述するが、スマートフォン100には、大きい声を出すとともに顔面の筋肉をしっかりと動かすように被訓練者10に促す画像が表示される。以下、スマートフォン100の詳細について説明する。

10

【0025】

[2.ハードウェア構成]

図2は、スマートフォン100のハードウェア構成の一例を示す図である。図2に示されるように、スマートフォン100は、カメラ130と、ディスプレイ140と、マイク150と、スピーカ160と、制御部170と、記憶部180と、通信モジュール190とを含んでいる。スマートフォン100に含まれる各構成要素は、バスを介して電氣的に接続されている。

【0026】

カメラ130は、被写体像を撮像し、画像データを生成するように構成されている。カメラ130は、たとえば、被訓練者10(図1)を撮像し、動画像データを生成する。カメラ130は、たとえば、CMOSイメージセンサ又はCCDイメージセンサ等のイメージセンサを含んでいる。

20

【0027】

ディスプレイ140は、画像を表示するように構成されている。ディスプレイ140は、たとえば、カメラ130によって生成された動画像データが示す動画像を表示する。ディスプレイ140は、たとえば、液晶ディスプレイ又は有機ELディスプレイ等のディスプレイによって構成される。

【0028】

マイク150は、マイク150の周囲の音に基づいて音声データを生成するように構成されている。マイク150は、たとえば、被訓練者10が発した声に基づいて音声データを生成する。

30

【0029】

スピーカ160は、音声データが示す音声を出力するように構成されている。スピーカ160は、たとえば、被訓練者10の声に基づいて生成された音声データが示す音声を出力する。

【0030】

制御部170は、CPU(Central Processing Unit)172、RAM(Random Access Memory)174及びROM(Read Only Memory)176等を含み、情報処理に応じて各構成要素の制御を行なうように構成されている。

40

【0031】

記憶部180は、たとえば、フラッシュメモリ等のメモリである。記憶部180は、たとえば、制御プログラム181を記憶するように構成されている。制御プログラム181は、制御部170によって実行されるスマートフォン100の制御プログラムである。制御部170が制御プログラム181を実行する場合に、制御プログラム181は、RAM174に展開される。そして、制御部170は、RAM174に展開された制御プログラム181をCPU172によって解釈及び実行することにより、各構成要素を制御する。

【0032】

通信モジュール190は、外部機器と通信するように構成されている。通信モジュール190は、たとえば、LTE(Long Term Evolution)モジュール、無線LANモジュール

50

ル等で構成される。

【 0 0 3 3 】

[3 . ソフトウェア構成]

図 3 は、制御部 1 7 0 によって実現される各ソフトウェアモジュールの関係の一例を示す図である。図 3 に示されるように、顔領域抽出部 1 3 1、画素移動量算出部 1 3 2、顔移動量補正部 1 3 3、筋活動量推定部 1 3 4、第 1 判定部 1 3 5、音声特徴抽出部 1 5 1 及び第 2 判定部 1 5 2 の各々は、ソフトウェアモジュールであり、制御部 1 7 0 が制御プログラム 1 8 1 を実行することによって実現されている。

【 0 0 3 4 】

顔領域抽出部 1 3 1 は、カメラ 1 3 0 によって生成された動画データに基づいて、被訓練者 1 0 の顔に対応する領域を抽出するように構成されている。顔領域の抽出方法としては、公知の種々の方法が用いられる。

10

【 0 0 3 5 】

画素移動量算出部 1 3 2 は、カメラ 1 3 0 によって生成された動画データに基づいて、各領域のオプティカルフローを算出するように構成されている。オプティカルフローの算出方法としては、公知の種々の方法が用いられる。ここで、各領域は、画像に含まれる各画素によって構成されてもよいし、画像に含まれる複数画素によって構成されてもよい。また、画素移動量算出部 1 3 2 は、領域毎に、算出されたオプティカルフローの大きさを示す画像を生成し、生成された画像をディスプレイ 1 4 0 に出力する。

【 0 0 3 6 】

顔移動量補正部 1 3 3 は、顔領域抽出部 1 3 1 によって抽出された顔領域の移動量及び移動方向を算出し、画素移動量算出部 1 3 2 によって算出されたオプティカルフローから減算するように構成されている。これにより、顔の移動量を差し引いた、顔面の筋肉の動きを示すオプティカルフローを算出することができる。

20

【 0 0 3 7 】

筋活動量推定部 1 3 4 は、各領域のオプティカルフローの大きさの和を算出することによって、被訓練者 1 0 の顔面の筋肉の動き量を推定するように構成されている。すなわち、筋活動量推定部 1 3 4 は、被訓練者 1 0 の口の動き量を推定するように構成されている。推定された顔面の筋肉の動き量（各領域のオプティカルフローの大きさの和）は、ディスプレイ 1 4 0 に出力される。

30

【 0 0 3 8 】

第 1 判定部 1 3 5 は、筋活動量推定部 1 3 4 によって推定された顔面の筋肉の動き量が第 1 所定量より小さい状態が所定時間継続したか否かを判定するように構成されている。第 1 所定量は、顔面の筋肉の動き量がこれよりも小さい場合に期待される発話訓練効果が得られない値である。顔面の筋肉の動き量が第 1 所定量よりも小さい状態が所定時間継続した場合に、第 1 警告画像がディスプレイ 1 4 0 に出力される。

【 0 0 3 9 】

音声特徴抽出部 1 5 1 は、マイク 1 5 0 によって生成された音声データに基づいて、被訓練者 1 0 が発した声の特徴量を抽出するように構成されている。音声特徴抽出部 1 5 1 は、たとえば、被訓練者 1 0 が発した声の大きさを抽出する。また、音声特徴抽出部 1 5 1 は、抽出された声の大きさを示す画像を生成し、生成された画像をディスプレイ 1 4 0 に出力する。

40

【 0 0 4 0 】

第 2 判定部 1 5 2 は、音声特徴抽出部 1 5 1 によって抽出された声の特徴量が第 2 所定量より小さい状態が所定時間継続したか否かを判定するように構成されている。第 2 所定量は、声の特徴量がこれよりも小さい場合に期待される発話訓練効果が得られない値である。声の特徴量が第 2 所定量よりも小さい状態が所定時間継続した場合に、第 2 警告画像がディスプレイ 1 4 0 に出力される。

【 0 0 4 1 】

[4 . 動作]

50

本実施の形態に従うスマートフォン100においては、制御部170によって、動画表示処理、オプティカルフロー表示処理、筋活動量表示処理、音声特徴量表示処理及び警告メッセージ表示処理が並列的に実行されている。以下、各処理について順に説明する。

【0042】

(4-1. 動画表示処理)

図4は、動画表示処理の実行手順を示すフローチャートである。このフローチャートに示される処理は、予め定められた周期で実行される。

【0043】

図4を参照して、制御部170は、被訓練者10の顔を含む動画像を撮像し動画像データを生成するとともに、被訓練者10の声を含む音声データを生成するようにカメラ130及びマイク150をそれぞれ制御する(ステップS100)。制御部170は、生成された動画像データに基づいて、動画に含まれる顔領域を抽出する(ステップS110)。制御部170は、被訓練者10に読ませる文章、抽出された顔領域を囲む枠、及び、動画像データが示す動画を重畳して表示するようにディスプレイ140を制御する(ステップS120)。被訓練者10に読ませる文章を示すテキストデータは、たとえば、記憶部180(図2)に予め記憶されている。

10

【0044】

図5は、ディスプレイ140に表示される画像の一例を示す図である。図5に示されるように、ディスプレイ140には、被訓練者10を含む動画、被訓練者10の顔領域を囲む顔枠200、及び、被訓練者10に読ませる文章210が表示されている。スマートフォン100によれば、被訓練者10に音読させる文章210がディスプレイ140に表示されるため、被訓練者10は、表示される文章を音読するだけで発話訓練を行なうことができる。

20

【0045】

(4-2. オプティカルフロー表示処理)

図6は、オプティカルフロー表示処理の実行手順を示すフローチャートである。このフローチャートに示される処理は、予め定められた周期で実行される。

【0046】

図6を参照して、制御部170は、動画表示処理において生成された動画像データに基づいて、各領域のオプティカルフローを算出する(ステップS200)。制御部170は、領域毎に、オプティカルフローの大きさ及び方向を示す画像を生成する(ステップS210)。制御部170は、生成された画像を動画に重畳表示するようにディスプレイ140を制御する(ステップS220)。

30

【0047】

再び図5を参照して、ディスプレイ140においては、被訓練者10の口が移動した軌跡を示す線240(オプティカルフロー)の画像が動画に重畳表示される。スマートフォン100によれば、被訓練者10の口の動きの量を示す画像が動画像に重畳して表示されるため、口の動きが不十分か否かを被訓練者10に視覚的に認識させることができる。その結果、スマートフォン100によれば、被訓練者10が口の周りの筋肉を大きく動かすことを意識して発話訓練を行なうことができるため、より効果的に被訓練者10の発話訓練を行なうことができる。

40

【0048】

(4-3. 筋活動量表示処理)

図7は、筋活動量表示処理の実行手順を示すフローチャートである。このフローチャートに示される処理は、予め定められた周期で実行される。

【0049】

図7を参照して、制御部170は、動画表示処理において生成された動画像データに基づいて、被訓練者10の顔領域を抽出するとともに、顔領域の動き(大きさ及び方向)を抽出する(ステップS300)。制御部170は、オプティカルフロー表示処理において算出されたオプティカルフローから、ステップS300において抽出された顔領域の動き

50

を減算することによって、オプティカルフローの補正を行なう（ステップS310）。制御部170は、各領域の補正後のオプティカルフローの大きさの和を算出することによって、被訓練者10の顔面の筋肉の動き量（以下、「筋活動量」とも称する。）を推定する（ステップS320）。制御部170は、推定された顔面の筋肉の動き量（各領域のオプティカルフローの大きさの和）を示す画像を生成し、該画像を表示するようにディスプレイ140を制御する（ステップS330）。

【0050】

再び図5を参照して、ディスプレイ140においては、レベルメータ220のような顔面の筋肉の動き量を示す画像が動画に重畳表示される。スマートフォン100によれば、被訓練者10の発話に関する評価結果（たとえば、口を含む顔面の筋肉の動き量）がディスプレイ140に表示されるため、被訓練者10は、評価結果を確認しながら発話訓練を行なうことができる。

10

【0051】

（4-4.音声特徴量表示処理）

図8は、音声特徴量表示処理の実行手順を示すフローチャートである。このフローチャートに示される処理は、予め定められた周期で実行される。

【0052】

図8を参照して、制御部170は、動画表示処理において生成された音声データに基づいて、被訓練者10の声の特徴量（たとえば、大きさ）を抽出する（ステップS400）。制御部170は、抽出された声の特徴量を示す画像を生成し、該画像を表示するようにディスプレイ140を制御する（ステップS410）。

20

【0053】

再び図5を参照して、ディスプレイ140においては、レベルメータ230のような声の特徴量を示す画像が動画に重畳表示される。スマートフォン100によれば、被訓練者10の発話に関する評価結果（たとえば、声の大きさ）がディスプレイ140に表示されるため、被訓練者10は、評価結果を確認しながら発話訓練を行なうことができる。

【0054】

（4-5.警告メッセージ表示処理）

図9は、警告メッセージ表示処理の実行手順を示すフローチャートである。このフローチャートに示される処理は、予め定められた周期で実行される。

30

【0055】

図9を参照して、制御部170は、筋活動量表示処理において推定された筋活動量が第1所定量よりも小さい状態が所定時間継続したか否かを判定する（ステップS500）。筋活動量が第1所定量以上であると判定されると（ステップS500においてNO）、処理はステップS510に移行する。一方、筋活動量が第1所定量よりも小さい状態が所定時間継続したと判定されると（ステップS500においてYES）、制御部170は、第1警告画像を表示するようにディスプレイ140を制御する（ステップS510）。

【0056】

図10は、ディスプレイ140に表示される画像の一例を示す図である。図10に示されるように、筋活動量が第1所定量よりも小さい状態が所定時間継続した場合には、第1警告画像250（「もっと口を動かして！」）がディスプレイ140に表示される。スマートフォン100によれば、被訓練者10の発話が所定要件を満たさない場合に第1警告画像250が表示されるため、被訓練者10は、自らの発話が所定要件を満たしていないことを視覚的に認識することができる。

40

【0057】

再び図9を参照して、次に、制御部170は、音声特徴量表示処理において抽出された音声特徴量が第2所定量よりも小さい状態が所定時間継続したか否かを判定する（ステップS520）。音声特徴量が第2所定量以上であると判定されると（ステップS520においてNO）、処理はステップS500に移行する。一方、音声特徴量が第2所定量よりも小さい状態が所定時間継続したと判定されると（ステップS520においてYES）、

50

制御部 170 は、第 2 警告画像を表示するようにディスプレイ 140 を制御する（ステップ S530）。

【0058】

再び図 10 を参照して、音声特徴量が第 2 所定量よりも小さい状態が所定時間継続した場合には、第 2 警告画像 260（「もっと大きな声で！」）がディスプレイ 140 に表示される。スマートフォン 100 によれば、被訓練者 10 の発話が所定要件を満たさない場合に第 2 警告画像 260 が表示されるため、被訓練者 10 は、自らの発話が所定要件を満たしていないことを視覚的に認識することができる。

【0059】

[5. 特徴]

以上のように、本実施の形態に従うスマートフォン 100 において、ディスプレイ 140 は、被訓練者 10 の口の動きの量を示す画像を動画像に重畳して表示する。スマートフォン 100 によれば、被訓練者 10 の口の動きの量を示す画像が動画像に重畳して表示されるため、口の動きが不十分か否かを被訓練者 10 に視覚的に認識させることができる。その結果、スマートフォン 100 によれば、被訓練者 10 が口の周りの筋肉を大きく動かすことを意識して発話訓練を行なうことができるため、より効果的に被訓練者 10 の発話訓練を行なうことができる。

【0060】

なお、スマートフォン 100 は、「発話訓練システム」の一例であり、カメラ 130 は、「撮像手段」の一例であり、ディスプレイ 140 は、「表示手段」の一例である。また、画素移動量算出部 132 は、「算出手段」及び「生成手段」の一例である。

【0061】

[6. 実験]

本発明者は、以下の実験を行なった。本実験は、防音室で行なわれた。実験に先立ち、実験参加者に実験の説明を行なった。次に、実験参加者に発話訓練の意義を説明し、意欲を持って実験に参加するよう依頼した。声量や話速は、高校の教室で朗読することをイメージするよう指示した。音声収録及び発話訓練は立位にて行なった。訓練時は、PC（Personal Computer）のディスプレイ（EIZO EV2450）を実験参加者の顔の正面にくるよう配置し、正面を向いた状態で練習できるようにした。なお、本実験においては、上記実施の形態に従うスマートフォン 100 において実装されたアプリケーションが PC にインストールされている。

【0062】

実験では、まず実験参加者の訓練前の音声を録音し、VAS（Visual analog scale）にてその発話がどの程度うまくできたかを自己評価させた。続いて、前歯で割り箸を噛んだ状態で、上記 PC（発話訓練システム）を用いて 3 分間練習した。その後、訓練後の音声を収録し、再び VAS を計測した。音声はコンデンサマイク（Sony ECM-77B）とレコーダ（Marantz PMD671）を用いて標準化周波数 16 kHz、量子化 16 bit にて収録した。

【0063】

図 11 は、訓練前後に録音した音声の振幅を示す図である。図 12 は、訓練前後に録音した音声の基本周波数の変化幅を示す図である。これらの結果は、各実験参加者の 14 文の平均値の分布を示している。図 11 及び図 12 に示されるように、上記発話訓練システムを用いた訓練によって振幅と基本周波数の変化幅がともに上昇する傾向にあることがわかる。訓練前後の中央値の比較では、振幅が 4.3 dB、基本周波数の変化幅が 1.19 semitone 上昇した。図は示さないが、基本周波数の平均値も訓練後に上昇する傾向にあった。

【0064】

図 13 は、訓練前後に計測した VAS を示す図である。図 13 に示されるように、ほぼ全ての実験参加者が、訓練によってうまく読めるようになったという自覚を持った。実験後には「ハキハキ言えるようになった」、「サ行、タ行が良くなった」、「（口の）横の筋肉が動きやすくなった」などの肯定的なコメントが多く聞かれた。顔面の動きをフィー

10

20

30

40

50

ドバックすることにより、わずか3分間の練習でも口の動きを改善する効果があったと考えられる。

【0065】

[7.変形例]

以上、実施の形態について説明したが、本発明は、上記実施の形態に限定されるものではなく、その趣旨を逸脱しない限りにおいて、種々の変更が可能である。以下、変形例について説明する。

【0066】

(7-1)

上記実施の形態においては、筋活動量を示す画像(口の動きの量を示す画像)として線240が用いられた。しかしながら、筋活動量を示す画像は、線240に限定されない。筋活動量を示す画像は、たとえば、動きの方向及び大きさを示す矢印であってもよい。また、筋活動量を示すために、たとえば、動き量が多い部分と動き量が少ない部分とで色を異ならせてもよい。たとえば、動き量が多い領域は赤色で表現し、動き量が少ない領域は青色で表現してもよい。

10

【0067】

(7-2)

上記実施の形態においては、カメラ130によって撮像された動画像全体のオブティカルフローが算出された。しかしながら、オブティカルフローが算出される範囲はこれに限定されない。たとえば、被訓練者10の顔領域のみのオブティカルフローが算出されてもよいし、被訓練者10の顔の下半分の領域のみのオブティカルフローが算出されてもよいし、被訓練者10の口領域のみのオブティカルフローが算出されてもよい。オブティカルフローを算出する領域を絞ることで、制御部170による計算量を減らすことができる。

20

【0068】

(7-3)

上記実施の形態においては、被訓練者10の画像及び音声を用いて発話訓練が行われた。しかしながら、被訓練者10の音声は、必ずしも発話訓練に用いられる必要はない。

【0069】

(7-4)

上記実施の形態においては、被訓練者10の口の動き量を得るためにオブティカルフローが算出された。しかしながら、必ずしもオブティカルフローが算出されなくてもよい。たとえば、単に動画像におけるフレーム間の差分を算出することによって、被訓練者10の口の動き量が得られてもよい。

30

【0070】

(7-5)

上記実施の形態においては、スマートフォンにおいて発話訓練システムが実現されたが、本発明に従う発話訓練システムは、たとえば、PC、タブレット等によって実現されてもよい。

【0071】

(7-6)

上記実施の形態において、発話訓練中に、ディスプレイ140に講師の手本動画があわせて表示されてもよい。

40

【0072】

(7-7)

上記実施の形態においては、被訓練者10の顔の領域毎のオブティカルフローが算出されている。したがって、たとえば、被訓練者10の顔の何れの領域の動きが不足しているかを算出することも可能である。たとえば、被訓練者10の顔の何れの領域の動きが不足しているかを示す警告画像がディスプレイ140に表示されてもよい。

【0073】

(7-8)

50

上記実施の形態において、たとえば、被訓練者10の発話訓練の履歴が順次記憶部180に記憶されてもよい。これにより、たとえば、被訓練者10が新たに発話訓練を行なった場合に、前回と比較してどの部分が改善されたか、どの部分が悪くなったか等を被訓練者10に知らせることができる。

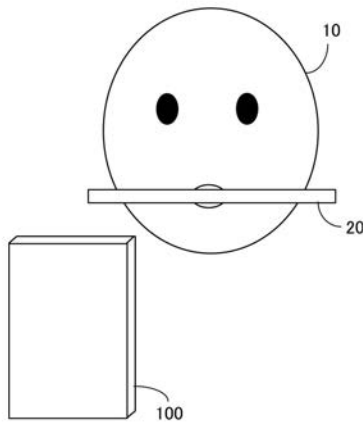
【符号の説明】

【0074】

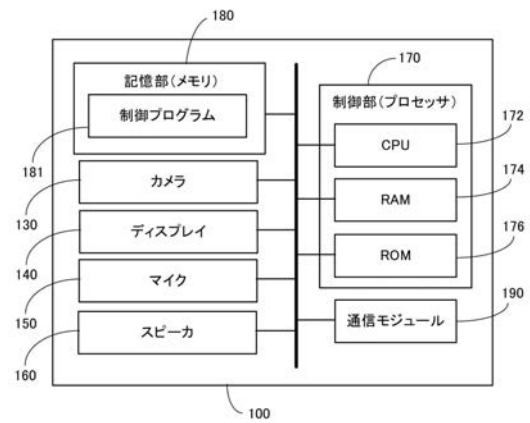
10 被訓練者、20 棒、100 スマートフォン、130 カメラ、131 顔領域抽出部、132 画素移動量算出部、133 顔移動量補正部、134 筋活動量推定部、135 第1判定部、140 ディスプレイ、150 マイク、151 音声特徴抽出部、152 第2判定部、160 スピーカ、170 制御部、172 CPU、174 RAM、176 ROM、180 記憶部、181 制御プログラム、190 通信モジュール、200 顔枠、210 文章、220、230 レベルメータ、240 線、250 第1警告画像、260 第2警告画像。

10

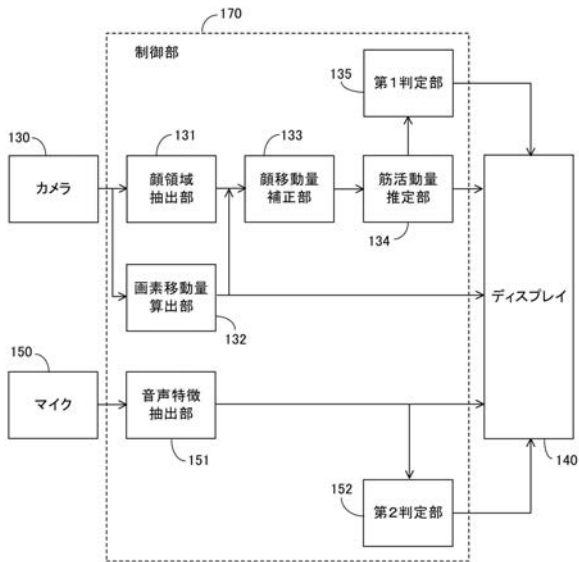
【図1】



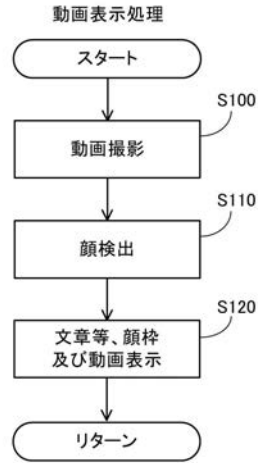
【図2】



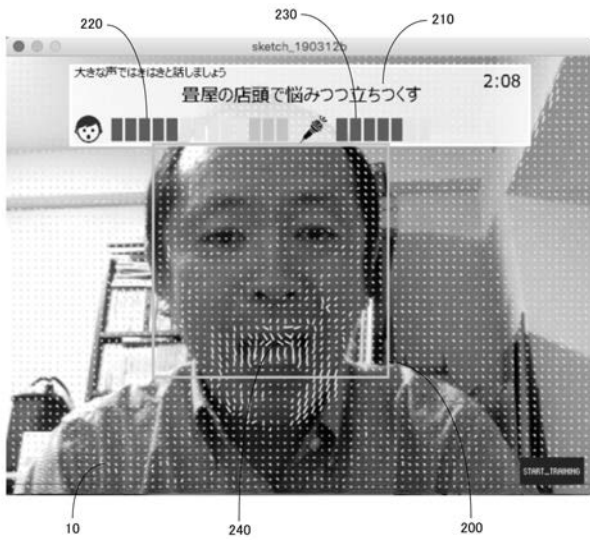
【図3】



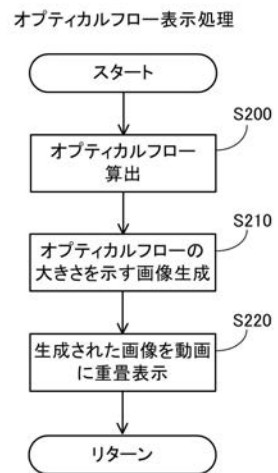
【図4】



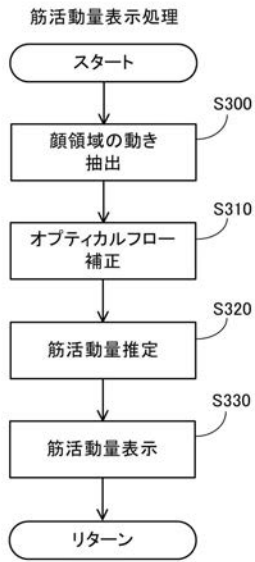
【図5】



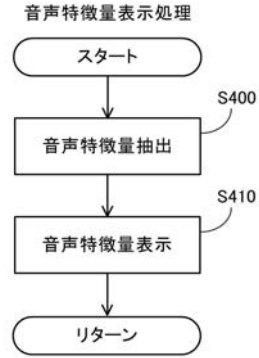
【図6】



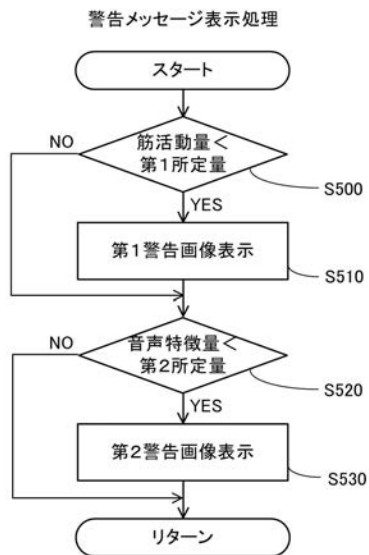
【 図 7 】



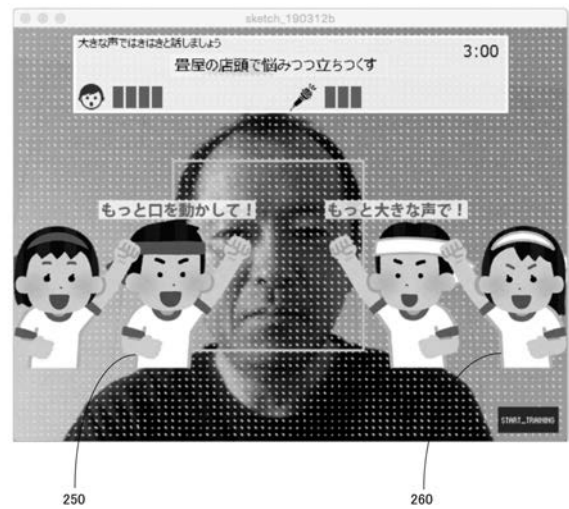
【 図 8 】



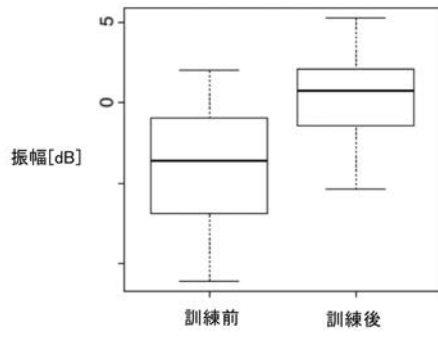
【 図 9 】



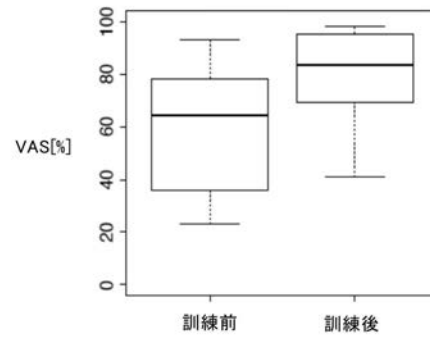
【 図 10 】



【 図 1 1 】



【 図 1 3 】



【 図 1 2 】

