

(19) 世界知的所有権機関
国際事務局



(43) 国際公開日
2003 年 9 月 25 日 (25.09.2003)

PCT

(10) 国際公開番号
WO 03/079328 A1

(51) 国際特許分類: G10L 15/00, H04N 5/278, G06F 17/28

区下野幌テクノパーク 1 丁目 1 番 1 4 号 Hokkaido (JP).

(21) 国際出願番号: PCT/JP03/03305

(22) 国際出願日: 2003 年 3 月 19 日 (19.03.2003)

(71) 出願人 および

(72) 発明者: 伊福部 達 (IFUKUBE, Tohru) [JP/JP]; 〒168-0063 東京都 杉並区 和泉 3-59-15 シティハウス 永福町 205 Tokyo (JP).

(25) 国際出願の言語: 日本語

(26) 国際公開の言語: 日本語

(74) 代理人: 橋爪 健 (HASHIZUME, Takeshi); 〒104-0061 東京都 中央区 銀座 3 丁目 1 3 番 1 7 号 Tokyo (JP).

(30) 優先権データ:

特願2002-77773 2002 年 3 月 20 日 (20.03.2002) JP

特願2003-68440 2003 年 3 月 13 日 (13.03.2003) JP

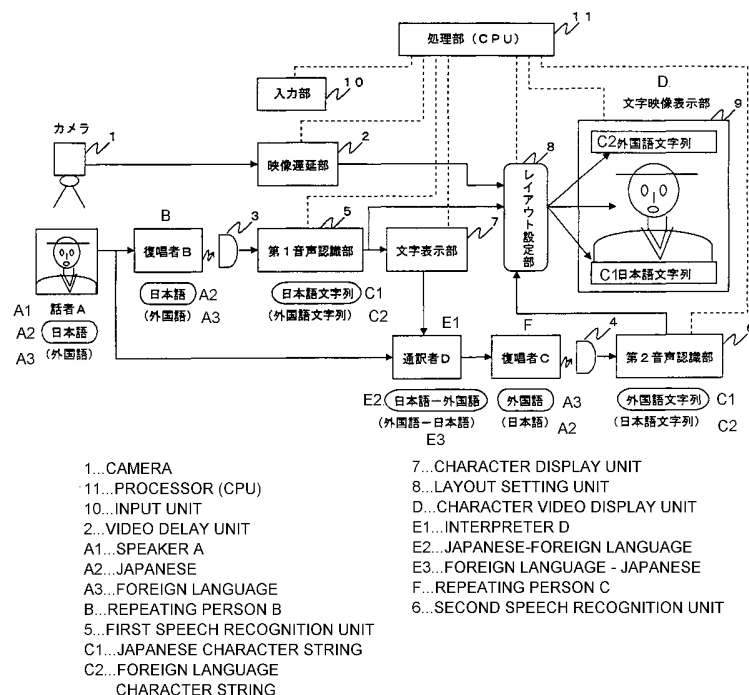
(71) 出願人 (米国を除く全ての指定国について): 科学技術振興事業団 (JAPAN SCIENCE AND TECHNOLOGY CORPORATION) [JP/JP]; 〒332-0012 埼玉県 川口市 本町 4-1-8 Saitama (JP). 株式会社ビー・ユー・ジー (B.U.G., INC.) [JP/JP]; 〒004-0015 北海道 札幌市 厚別

(81) 指定国 (国内): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

[続葉有]

(54) Title: AUDIO VIDEO CONVERSION APPARATUS AND METHOD, AND AUDIO VIDEO CONVERSION PROGRAM

(54) 発明の名称: 音声映像変換装置及び方法、音声映像変換プログラム



(57) Abstract: Speech of a speaker is repeated by a repeating person whose speech is recognized and a video of the speaker is delayed when displayed so that it is displayed together with characters, so that the speech of the speaker can easily be understood. A video delay unit (2) outputs delayed video data of video input to a camera (1) and delayed. A first speech recognition unit (5) recognizes the content of a first language of

[続葉有]



(84) 指定国 (広域): ARIPO 特許 (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシア特許 (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ特許 (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI 特許 (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

添付公開書類:

- 国際調査報告書
- 不利にならない開示又は新規性喪失の例外に関する申立て

2文字コード及び他の略語については、定期発行される各PCTガゼットの巻頭に掲載されている「コードと略語のガイダンスノート」を参照。

規則4.17に規定する申立て:

- すべての指定国のための不利にならない開示又は新規性喪失の例外に関する申立て (規則4.17(v))

a first repeating person input to a first speech input unit (3) and converts it into visible language data. A second speech recognition unit (6) recognizes the content of a second language of a second repeating person input to a second speech input unit (4) and converts it into second visible language data. A layout setting unit (8) receives the first and the second language data from the first and the second speech recognition unit (5, 6) and delayed video data from the video delay unit (2), sets a display layout of these data, creates a display video, and displays it on a character video display unit (9).

(57) 要約: 話者の声を復唱者が復唱した音声を認識し、話者の映像を遅らせて文字とともに表示することにより、話者の話したことを理解しやすくする。映像遅延部2は、カメラ1に入力された映像を遅延した遅延映像データを出力する。第1音声認識部5は、第1音声入力部3に入力された第1復唱者による第1言語の内容を認識して第1可視言語データに変換する。第2音声認識部6は、第2音声入力部4に入力された第2復唱者による第2言語の内容を認識して第2可視言語データに変換する。レイアウト設定部8は、第1及び第2音声認識部5、6からの第1及び第2言語データ及び映像遅延部2からの遅延映像データを入力し、それらデータの表示レイアウトを設定し、表示映像を生成し、文字映像表示部9に表示する。

明 細 書

音声映像変換装置及び方法、音声映像変換プログラム

5 技術分野

本発明は、音声映像変換装置及び方法、音声映像変換プログラムに係る。

背景技術

10 従来、聴覚障害者が参加しうる会議の支援手段としては、例えば、字幕放送や要約筆記がある。

一方、コンピュータによる音声自動認識技術は現時点では、使用前にあらかじめ使用者の音声で幾つかの単語や文章を読み上げて音声認識装置に入力し、使用者の音声の特徴を辞書に登録するという手法をとる。このように話者の音声を登録し、話題を限ったとしても、最高の認識率はせいぜい95%程度である。

本発明に抵触する論文等の報告は本発明者は発見していないが、NHKが放送映像に字幕を付ける際に復唱者による音声認識方式を採用している。また、プレスリリース（2003年1月20日）により、（株）ダイキン工業が「音声認識によるノンリニア文字おこしソフト（mospy）を新発売」との記事が発表される。これは映像と音声を一時停止と再生を繰り返しながら復唱し、音声認識装置を介して文字化するソフトウェアである。

25 発明の開示

しかしながら、このような従来の字幕放送や要約筆記については、異言語対応になっていないこと、字幕作りや要約作りには熟練を要すること、また、その熟練者が少ないことなど、普及に向けて大きな障壁があった。

一方、通常の音声自動認識技術に関しては、現在のところ誰の声でも正しく認識する不特定話者の音声認識は精度が極めて低く、雑音の多い環境下では使用できない場合が想定される。また、音声の認識時間が1秒ほどかかり、また、通訳者を介するとさらに2～3秒かかってしまう。

- 5 よって、音声認識された結果である文字列と話者の表情等に大きな時間差が生じ、そのため、話し手の唇の動きや表情、さらに手話などの視覚データを文意の理解に利用できなくなる。さらに、日本語の場合は、多くの同音意義語の漢字があるため前後関係から文意を推定できないと誤変換してしまう。現在の技術では文意を人工的に把握することは困難で
- 10 あり、漢字の選択は音声認識装置の使用者に任せている。また、現在の音声認識技術では、話者や話題が変わると途端に認識率が低下する。使用環境も静かなところに限られ、しかも、マイクロホンも特定のものを使用し、いつも口元の同じ位置に設置しなければならない。

- このように、従来は、音声認識装置を聴覚障害者のための会議支援や
- 15 通訳支援に利用することはむずかしかった。

さらに、上述のNHK方式及びダイキン工業社製品にはインターネット等の電気通信回路を利用していないので、遠隔地や在宅地にいる通訳者、復唱者によりユーザを支援するサービスを提供することはできなかった。

20

- 本発明は、以上の点に鑑み、不特定話者の声を復唱者が自己の声に変換し音声認識装置を介して文字に変換するとともに、話者の表情などの映像を遅らせて文字とともにスクリーンなどに表示することにより、聴覚障害者などが話者の話したことを理解しやすくするための音声映像変
- 25 換装置及び方法、音声映像変換プログラムを提供することを目的とする。

また、本発明は、聴覚障害者が出席するような国際会議、多国間・二国間会議等の会議において、講演者あるいは通訳者の音声を復唱者が復唱して音声認識装置に入力し、その結果である文字列を講演者の映像とともにスクリーンに表示するようにした会議支援のための音声映像変換

装置及び方法、音声映像変換プログラムを提供することを目的とする。

さらに、本発明は、異種言語を使って行われる国際会議等の通訳および会議の即時印刷（情報補償）、聴覚障害者等が参加する会議や授業の支援、電話から復唱者へ声を転送し文字化された情報をユーザに提供することを目的とする。また、本発明は、話者とユーザとの異なる言語体系間におけるコミュニケーションの補助を行うための音声映像変換装置及び方法、音声映像変換プログラムを提供することを目的とする。

また、本発明は、さらにインターネットなどの電気通信回線を用いて通信を行う電気通信回路により、話者の声と映像を遠隔地や在宅地にいる通訳者、復唱者及び修正者に転送する手段を付加することにより、ユーザがどこにいても本システムを利用できるようにすることを目的とする。本発明は、介在する復唱者や通訳者が在宅ビジネスとして利用すること、さらに、外出の難しい在宅の障害者が復唱者になることにより就労を支援することを目的とする。

本発明の第 1 の解決手段によると、

話者の表情映像を撮影するカメラと、

前記カメラにより撮影された映像信号に対して予め設定された遅延時間差を与え、遅延映像データを出力する映像遅延部と、

話者が話す第 1 言語の内容を復唱する第 1 復唱者の第 1 言語の内容が入力される第 1 音声入力部と、

話者が話す第 1 言語の内容を通訳した通訳者の第 2 言語の内容をさらに復唱する第 2 復唱者の第 2 言語の内容が入力される第 2 音声入力部と、

それぞれ、前記第 1 及び第 2 音声入力部から入力された第 1 及び第 2 言語の内容を認識して第 1 及び第 2 可視言語データに変換して出力する第 1 及び第 2 音声認識部と、

前記第 1 及び第 2 音声認識部から出力された第 1 及び第 2 可視言語データと、前記映像遅延部により遅延された話者の遅延映像データとを入力し、表示状態を設定してこれらデータを同期又は略同期させた表示映

像を生成するレイアウト設定部と、

前記レイアウト設定部からの出力に従い、第1及び第2可視言語データと遅延映像データとを同期又は略同期させた表示映像を表示する文字映像表示部と

- 5 前記第1及び第2音声認識部、前記映像遅延部、前記レイアウト設定部のいずれか又は複数の各部の各種設定を行うための入力部と、

前記第1及び第2音声認識部、前記映像遅延部、前記入力部、前記レイアウト設定部の各部を制御する処理部と
を備えた音声映像変換装置が提供される。

10

本発明の第2の解決手段によると、

話者の表情映像を撮影するカメラと、

前記カメラにより撮影された映像信号に対して予め設定された遅延時間差を与え、遅延映像データを出力する映像遅延部と、

- 15 話者又は通訳者が話す第1言語の内容を復唱する第1復唱者の第1言語の内容が入力される第1音声入力部と、

第1音声認識部は、前記第1音声入力部から入力された第1言語の内容を認識して第1可視言語データに変換して出力する第1音声認識部と、

- 前記第1音声認識部から出力された第1可視言語データと、前記映像
20 遅延部により遅延された話者の遅延映像データとを入力し、表示状態を設定してこれらデータを同期又は略同期させた表示映像を生成するレイアウト設定部と、

前記レイアウト設定部からの出力に従い、第1可視言語データと遅延映像データとを同期又は略同期させた表示映像を表示する文字映像表示

25 部と

前記第1音声認識部、前記映像遅延部、前記レイアウト設定部のいずれか又は複数の各部の各種設定を行うための入力部と、

前記第1音声認識部、前記映像遅延部、前記入力部、前記レイアウト設定部の各部を制御する処理部と

を備えた音声映像変換装置が提供される。

本発明の第3の解決手段によると、

話者の音声を可視言語データに変換して話者の映像データとともに表示
5 示するための音声映像変換方法又はプログラムであって、

処理部は、入力部からの指令又は適宜の記憶部により予め定められた
設定に従い、第1及び第2音声認識部及び映像遅延部の設定を行うステ
ップと、

処理部は、入力部からの指令又は適宜の記憶部により予め定められた
10 設定に従い、レイアウト設定部の設定を行うステップと、

カメラは、話者の映像を入力するステップと、

映像遅延部は、処理部による設定及び制御に従い、カメラに入力され
た映像を遅延及び必要に応じて適宜の画像処理を行い、遅延映像デー
タを出力するステップと、

15 第1音声入力部は、話者による第1言語の内容を復唱する第1復唱者
による第1言語の内容を入力するステップと、

第1音声認識部は、第1音声入力部に入力された第1復唱者による第
1言語の内容を認識して第1可視言語データに変換するステップと、

第2音声入力部は、話者による第1言語の内容を通訳者が通訳した第
20 2言語の内容を第2復唱者が復唱し、その復唱された第2言語の内容を
入力するステップと、

第2音声認識部は、第2音声入力部に入力された第2復唱者による第
2言語の内容を認識して第2可視言語データに変換するステップと、

レイアウト設定部は、処理部による設定及び制御に従い、第1及び第
25 2音声認識部からの第1及び第2言語データ及び映像遅延部からの遅延
映像データを入力し、それらデータの表示レイアウトを設定し、画像処
理によりこれらデータを同期又は略同期させた表示映像を生成及び出力
するステップと、

文字映像表示部は、レイアウト設定部からの出力に従い、第1及び第

2 言語データ及び映像遅延データを同期又は略同期させた表示映像を表示するステップと、

を含む音声映像変換方法、及び、これら各ステップをコンピュータに実行させるためのプログラムが提供される。

5

本発明の第4の解決手段によると

話者の音声を可視言語データに変換して話者の映像データとともに表示するための音声映像変換又はプログラムであって、

10 処理部は、入力部からの指令又は適宜の記憶部により予め定められた設定に従い、第1音声認識部及び映像遅延部の設定を行うステップと、

処理部は、入力部からの指令又は適宜の記憶部により予め定められた設定に従い、レイアウト設定部の設定を行うステップと、

カメラは、話者の映像を入力するステップと、

15 映像遅延部は、処理部による設定及び制御に従い、カメラに入力された映像を遅延及び必要に応じて適宜の画像処理を行い、遅延映像データを出力するステップと、

第1音声入力部は、話者又は通訳者による第1言語の内容を復唱する第1復唱者による第1言語の内容を入力するステップと、

20 第1音声認識部は、第1音声入力部に入力された第1復唱者による第1言語の内容を認識して第1可視言語データに変換するステップと、

レイアウト設定部は、処理部による設定及び制御に従い、第1音声認識部からの第1言語データ及び映像遅延部からの遅延映像データを入力し、それらデータの表示レイアウトを設定し、画像処理によりこれらデータを同期又は略同期させた表示映像を生成及び出力するステップと、

25 文字映像表示部は、レイアウト設定部からの出力に従い、第1言語データ及び映像遅延データを同期又は略同期させた表示映像を表示するステップと、

を含む音声映像変換方法、及び、これら各ステップをコンピュータに実行させるためのプログラムが提供される。

本発明の第 5 の解決手段によると

話者が話す第 1 言語の内容を復唱する第 1 復唱者の第 1 言語の内容を認識して第 1 可視言語データに変換して出力する第 1 音声認識部と、前
5 記第 1 音声認識部の各種設定を行うための第 1 入力部と、前記第 1 音声認識部及び前記第 1 入力部を制御する第 1 処理部とを有する第 1 認識装置と、

話者が話す第 1 言語の内容を通訳した通訳者の第 2 言語の内容をさらに復唱する第 2 復唱者の第 2 言語の内容を認識して第 2 可視言語データ
10 に変換して出力する第 2 音声認識部と、前記第 2 音声認識部の各種設定を行うための第 2 入力部と、前記第 2 音声認識部及び前記第 2 入力部を制御する第 2 処理部とを有する第 2 認識装置と、

前記第 1 及び第 2 認識装置からの出力が入力され、文字及び映像を表示するための表示装置と

15 を備え、

前記表示装置は、

カメラにより撮影された映像信号に対して予め設定された遅延時間差を与え、遅延映像データを出力する映像遅延部と、

前記第 1 及び第 2 認識装置から出力された第 1 及び第 2 可視言語データと、前記映像遅延部により遅延された話者の遅延映像データとを入力し、表示状態を設定してこれらデータを同期又は略同期させた表示映像を生成するレイアウト設定部と、

前記レイアウト設定部から出力された表示映像を表示する文字映像表示部と

25 前記映像遅延部及び前記レイアウト設定部の各種設定を行うための第 3 入力部と、

前記映像遅延部、前記第 3 入力部、前記レイアウト設定部の各部を制御する第 3 処理部と

を有する音声映像変換装置が提供される。

本発明の第 6 の解決手段によると

話者又は通訳者が話す第 1 言語の内容を復唱する第 1 復唱者の第 1 言語の内容を認識して第 1 可視言語データに変換して出力する第 1 音声認識部と、前記第 1 音声認識部の各種設定を行うための第 1 入力部と、前
5 記第 1 音声認識部及び前記第 1 入力部を制御する第 1 処理部とを有する第 1 認識装置と、

前記第 1 認識装置からの出力が入力され、文字及び映像を表示するための表示装置と
を備え、

10 前記表示装置は、

カメラにより撮影された映像信号に対して予め設定された遅延時間差を与え、遅延映像データを出力する映像遅延部と、

前記第 1 認識装置から出力された第 1 可視言語データと、前記映像遅延部により遅延された話者の遅延映像データとを入力し、表示状態を設定してこれらデータを同期又は略同期させた表示映像を生成するレイア
15 ウト設定部と、

前記レイアウト設定部から出力された表示映像を表示する文字映像表示部と

前記映像遅延部及び前記レイアウト設定部の各種設定を行うための第
20 3 入力部と、

前記映像遅延部、前記第 3 入力部、前記レイアウト設定部の各部を制御する第 3 処理部と
を有する音声映像変換装置が提供される。

本発明の第 7 の解決手段によると

25 話者の音声を可視言語データに変換して話者の映像データとともに表示するための音声映像変換方法であって、

第 1 及び第 2 処理部、第 3 処理部は、それぞれ、第 1 及び第 2 入力部、第 3 入力部からの指令又は適宜の記憶部により予め定められた設定に従い、第 1 及び第 2 音声認識部、映像遅延部の設定を行うステップと、

第3処理部は、第3入力部からの指令又は適宜の記憶部により予め定められた設定に従い、レイアウト設定部の設定を行うステップと、

映像遅延部は、第3処理部による設定及び制御に従い、カメラに入力された話者の映像を遅延及び必要に応じて適宜の画像処理を行い、遅延
5 映像データを出力するステップと、

第1音声認識部は、話者による第1言語の内容を復唱する第1復唱者による第1言語の内容を認識して第1可視言語データに変換するステップと、

第2音声認識部は、話者による第1言語の内容を通訳者が通訳した第
10 2言語の内容を復唱した第2復唱者による第2言語の内容を認識して第2可視言語データに変換するステップと、

レイアウト設定部は、第3処理部による設定及び制御に従い、第1及び第2音声認識部からの第1及び第2可視言語データ及び映像遅延部からの遅延映像データを入力し、それらデータの表示レイアウトを設定し、
15 画像処理によりこれらデータを同期又は略同期させた表示映像を生成及び出力するステップと、

文字映像表示部は、レイアウト設定部からの出力に従い、第1及び第2可視言語データ及び映像遅延データを同期又は略同期させた表示映像を表示するステップと、

20 を含む音声映像変換方法が提供される。

本発明の第8の解決手段によると

話者の音声を可視言語データに変換して話者の映像データとともに表示するための音声映像変換方法であって、

第1及び第3処理部は、それぞれ、第1及び第3入力部からの指令又は適宜の記憶部により予め定められた設定に従い、第1音声認識部及び映像遅延部の設定を行うステップと、
25

第3処理部は、第3入力部からの指令又は適宜の記憶部により予め定められた設定に従い、レイアウト設定部の設定を行うステップと、

映像遅延部は、第3処理部による設定及び制御に従い、カメラに入力

された話者の映像を遅延及び必要に応じて適宜の画像処理を行い、遅延映像データを出力するステップと、

第1音声認識部は、話者又は通訳者による第1言語の内容を復唱する第1復唱者による第1言語の内容を認識して第1可視言語データに変換するステップと、

レイアウト設定部は、第3処理部による設定及び制御に従い、第1音声認識部からの第1可視言語データ及び映像遅延部からの遅延映像データを入力し、それらデータの表示レイアウトを設定し、画像処理によりこれらデータを同期又は略同期させた表示映像を生成及び出力するステップと、

文字映像表示部は、レイアウト設定部からの出力に従い、第1可視言語データ及び映像遅延データを同期又は略同期させた表示映像を表示するステップと、

を含む音声映像変換方法が提供される。

図面の簡単な説明

図1は、音声映像変換装置の第1の実施の形態の概略構成図。

図2は、処理部による音声変換処理の第1の実施の形態のフローチャート。

図3は、音声映像変換装置の第2の実施の形態の概略構成図。

図4は、処理部による音声変換処理の第2の実施の形態のフローチャート。

図5は、音声映像変換装置の第3の実施の形態の概略構成図。

図6は、音声映像変換装置の第4の実施の形態の概略構成図。

発明を実施するための最良の形態

以下、図面を用いて本発明の実施の形態を詳細に説明する。

1. 第1の実施の形態

図 1 は、音声映像変換装置の第 1 の実施の形態の概略構成図である。

本実施の形態は、特に、国際会議、多国間会議、二カ国間会議等の複数言語が関与する会議・会合・講義・授業・教育等におけるコミュニケーションを支援する。本実施の形態の音声映像変換装置は、カメラ 1、
5 映像遅延部 2、第 1 及び第 2 音声入力部 3、4、第 1 及び第 2 音声認識部 5、6、文字表示部 7、レイアウト設定部 8、文字映像表示部 9、入力部 10、処理部 11 を備える。

カメラ 1 は、話者 A の表情映像を撮影する。映像遅延部 2 は、カメラ
10 1 からの映像信号に対して予め設定された遅延時間差を与え、遅延映像データを出力する。映像遅延部 2 は、話者の表情映像を認識された文字と一緒に表示して、受け手の言語理解の補助となるようにするために所定の映像遅延時間を与える。この映像遅延時間は、聴覚障害者等の会議参加者の読話能力、話者 A・復唱者 B 又は C・通訳者 D の話すスピード
15 や能力等に応じて、適宜変更することができる。また、映像遅延部 2 は、話者 A の表情などの映像を拡大・縮小等の適宜の画像処理を行うようにしてもよい。

第 1 音声入力部 3 は、マイクロフォン等で構成され、話者 A の音声を
20 聞き取った特定の第 1 復唱者 B の音声による内容が入力される。一方、第 2 音声入力部 4 は、話者 A が話す内容を通訳者 D が通訳し、その通訳者 D の音声を聞き取った特定の第 2 復唱者 C の音声による内容が入力される。復唱者 B 又は C は、会議内に設けた静かな場所で、説話マイクロホン等の第 1 又は第 2 音声入力部 3、4 を通して音声入力することで、
25 環境雑音やマイクロホンの影響を解決することもできる。

第 1 及び第 2 音声認識部 5、6 は、それぞれ第 1 及び第 2 音声入力部 3、4 から入力された音声を認識して文字データ、表意データ等の第 1 及び第 2 可視言語データに変換して出力する。この例では、第 1 音声認

識部 5 は、話者 A が話す第 1 言語（例：日本語）を聞いた第 1 復唱者 B
により第 1 言語で復唱された内容が入力され、第 1 言語の可視言語デー
タ（例：日本語文字列）を出力する。一方、第 2 音声認識部 6 は、話者
A が話す第 1 言語（例：日本語）を聞いた通訳者 D が第 2 言語（例：英
5 語などの外国語）に通訳して、さらに、通訳者 D が話す第 2 言語を聞い
た第 2 復唱者 C により第 2 言語で復唱された内容が入力され、第 2 言語
の可視言語データ（例：英語などの外国語文字列）を出力する。

第 1 及び／又は第 2 音声認識部 5、6 は、音声を手 1 復唱者 B が復唱
10 した音声、通訳者 D の音声を第 2 復唱者 C が復唱した音声のいずれか又
は両方を選択できるようにしてもよい。第 1 及び／又は第 2 音声認識部
5、6 は、復唱者の音声を認識するように設定されており、話者 A が話
す話題又は会議の内容等により、第 1 及び／又は第 2 復唱者 B、C が第
1 及び／又は第 2 音声認識装置 5、6 に登録されている言語データベー
15 スを選択できる選択部を備えるようにしてもよい。

さらに、第 1 及び／又は第 2 音声認識部 5、6 は、仮名－漢字変換に
おいて誤変換される確率を計算する誤変換確率計算部と、誤変換確率計
算部で計算された確率に応じて漢字出力か仮名文字出力かを決定する出
20 力決定部を備えるようにしてもよい。第 1 及び／又は第 2 音声認識部 5、
6 は、日本語の同音意義語の漢字処理に関しては、誤認識の確率を音声
認識前に計算し、その確率が高い場合には仮名文字で表示するようにす
ることもできる。また、第 1 及び／又は第 2 音声認識部 5、6 に登録さ
れていない言葉は、第 1 及び／又は第 2 復唱者 B、C の判断により仮名
25 文字で表示するようにしてもよい。

文字表示部 7 は、第 1 音声認識部 5 により出力された第 1 言語の可視
言語データを可視表示する。通訳者 D は、文字表示部 7 により表示され
た第 1 可視言語データを見て通訳するようにしても良い。

レイアウト設定部 8 は、第 1 及び第 2 音声認識部 5、6 により認識された結果として出力された第 1 及び第 2 可視言語データと、映像遅延部 2 により遅延された話者 A の遅延映像データとを入力し、文字映像表示部 9 への表示状態を設定する。処理部 11 は、文字映像表示部 9 に表示される第 1 及び第 2 可視言語データ（文字データ）及び遅延映像データについての、単位時間当たりの行数、単位時間当たりの文字数、1 行当たりの文字数、色、大きさ、表示位置、その他の表示形式のいずれか又は複数を設定し、レイアウト設定部 8 は、処理部 11 による設定に応じて第 1 及び第 2 可視言語データ及び遅延映像データについての拡大・縮小等の適宜の画像処理を実行し、表示画像を生成する。

文字映像表示部 9 は、レイアウト設定部 8 により設定及び生成された出力に従い、第 1 及び第 2 音声認識部 5、6 により認識された結果として出力された第 1 及び第 2 可視言語データと、映像遅延部 2 により遅延された話者 A の遅延映像データとを組み合わせ表示する。

入力部 10 は、第 1 及び第 2 音声認識部 5、6、映像遅延部 2、レイアウト設定部 8 等の各部の各種設定、適宜のデータベースやメモリ等へのデータ入力指示を行う。処理部 11 は、小型コンピュータであって、第 1 及び第 2 音声認識部 5、6、映像遅延部 2、入力部 10、レイアウト設定部 8 等の各部を制御する。

図 2 に、処理部による音声変換処理の第 1 の実施の形態フローチャートを示す。

処理部 11 は、入力部 10 からの指令又は適宜の記憶部により予め定められた設定に従い、第 1、第 2 音声認識部 5、6 及び映像遅延部 2 の設定を行う（S01）。第 1、第 2 音声認識部 5、6 の設定では、例えば、漢字誤認識率の閾値、使用する言語データベース等を設定する。映像遅延部 2 の設定では、例えば、話者画像の遅延時間の設定又は選定を行う。

さらに、処理部 11 は、入力部 10 からの指令又は適宜の記憶部により
予め定められた設定に従い、レイアウト設定部 8 の設定を行う（S 03）。
レイアウト設定部 8 の設定では、文字映像表示部 9 に表示される第 1、
第 2 可視言語データ及び遅延映像データの表示状態・レイアウトを設定
5 する。可視言語データについては、例えば、呈示文字列数、呈示文字の
大きさ・フォント・色、文字列の表示位置、また、遅延映像データについ
ては、話者画像の大きさ、表示位置等がそれぞれ適宜設定される。

カメラ 1 は、話者 A の映像を入力する（S 05）。映像遅延部 2 は、処
10 理部 11 による設定及び制御に従い、カメラ 1 に入力された映像を遅延
及び必要に応じて適宜の画像処理を行い、遅延映像データを出力する（S
07）。

第 1 音声入力部 3 は、第 1 復唱者 B による音声を入力する（S 11）。
15 第 1 音声認識部 5 は、処理部 11 による設定及び制御に従い、第 1 音声
入力部 3 に入力された第 1 復唱者 B による第 1 言語を認識して第 1 可視
言語データ（例：日本語文字列）に変換する（S 13）。さらに、必要に
応じて、文字表示部 7 は、第 1 音声認識部 5 から出力された第 1 可視言
語データを表示する（S 15）。

20 第 2 音声入力部 4 は、通訳者 D が、話者音声及び／又は文字表示部 7
に表示された第 1 可視言語データに基づき通訳した音声を第 2 復唱者 C
が復唱し、その復唱された音声を入力する（S 17）。第 2 音声認識部 6
は、処理部 11 による設定及び制御に従い、第 2 音声入力部 4 に入力さ
25 れた第 2 復唱者 C による第 2 言語を認識して第 2 可視言語データ（例：
外国語文字列）に変換する（S 19）。

レイアウト設定部 8 は、処理部 11 による設定及び制御に従い、第 1
及び第 2 音声認識部 5、6 からの第 1 及び第 2 可視言語データ及び映像

遅延部 2 から遅延映像データを入力し、それらデータの表示レイアウトを設定し、必要に応じて適宜の画像処理により表示画像を生成及び出力する（S 2 1）。文字映像表示部 9 は、レイアウト設定部 8 からの出力に従い、第 1 及び第 2 可視言語データ及び映像遅延部 2 を適宜表示する（S 2 3）。

処理部 1 1 は、設定変更があるときはステップ S 0 1 に戻り処理を実行する（S 2 5）。また、処理部 1 1 は、設定変更が無い場合、話者 A 変更が無いときはステップ S 0 3 の後の処理に移り、一方、話者 A 変更があるときは、処理を終了して（S 2 7）、改めて処理を実行することができる。

2. 第 2 の実施の形態

図 3 は、音声映像変換装置の第 2 の実施の形態の概略構成図である。

本実施の形態は、特に、国内会議、二カ国間会議等の会議・会合・講義・授業・教育等におけるコミュニケーションを支援する。本実施の形態の音声映像変換装置は、カメラ 1、映像遅延部 2、第 1 及び第 2 音声入力部 3、4、第 1 音声認識部 5、文字表示部 7、レイアウト設定部 8、文字映像表示部 9、入力部 1 0、処理部 1 1、及び、選択部 2 0 を備える。

第 1 の実施の形態と比較すると、第 2 音声認識部が省略され、選択部 2 0 がさらに備えられた点が、異なるが、他の構成及び動作は同様である。なお、第 2 音声入力部及び選択部 2 0 は、必要に応じて、さらに省略してもよい。

図 4 に、処理部による音声変換処理の第 2 の実施の形態フローチャートを示す。

第 1 の実施の形態と比較すると、主に、ステップ S 1 7 ～ S 1 9 が省

略された点が異なる。また、第1音声入力部3には、話者の音声を復唱した復唱者Bの音声か、話者の音声を通訳した通訳者Dの音声を復唱者Cが復唱した音声かのいずれかが入力される。

5 処理部11は、入力部10からの指令又は適宜の記憶部により予め定められた設定に従い、第1音声認識部5及び映像遅延部2及び選択部20の設定を行う(S101)。なお、選択部20が省略されているときは、その設定は不要である。第1音声認識部5の設定では、例えば、漢字誤認識率の閾値、使用する言語データベース等を設定する。映像遅延部2
10 の設定では、例えば、話者画像の遅延時間の設定又は選定を行う。さらに、処理部11は、入力部10からの指令又は適宜の記憶部により予め定められた設定に従い、レイアウト設定部8の設定を行う(S103)。レイアウト設定部8の設定では、文字映像表示部9に表示される第1可視言語データ(この例では、日本語文字列又は外国語文字列)及び遅延
15 映像データの表示状態・レイアウトを設定する。可視言語データについては、例えば、呈示文字列数、呈示文字の大きさ・フォント・色、文字列の表示位置、また、遅延映像データについては、話者画像の大きさ、表示位置等がそれぞれ適宜設定される。

20 カメラ1は、話者Aの映像を入力する(S105)。映像遅延部2は、処理部11による設定及び制御に従い、カメラ1に入力された映像を遅延及び必要に応じて適宜の画像処理を行い、遅延映像データを出力する(S107)。

25 第1音声入力部3は、第1復唱者B又は第2復唱者Cによる音声を入力する(S111)。第1音声認識部5は、処理部11による設定及び制御に従い、第1音声入力部3に入力された第1復唱者B又は第2復唱者Cによる第1言語(この例では、日本語又は外国語)を認識して第1可視言語データ(この例では、日本語文字列又は外国語文字列)に変換す

る（S 1 1 3）。さらに、必要に応じて、文字表示部 7 は、第 1 音声認識部 5 から出力された第 1 可視言語データを表示する（S 1 1 5）。

レイアウト設定部 8 は、処理部 1 1 による設定及び制御に従い、第 1 音声認識部 5 からの第 1 可視言語データ及び映像遅延部 2 から遅延映像データを入力し、それらデータの表示レイアウトを設定し、必要に応じて適宜の画像処理により表示画像を生成及び出力する（S 1 2 1）。文字映像表示部 9 は、レイアウト設定部 8 からの出力に従い、第 1 可視言語データ及び映像遅延部 2 を適宜表示する（S 1 2 3）。

処理部 1 1 は、設定変更があるときはステップ S 1 0 1 に戻り処理を実行する（S 1 2 5）。また、処理部 1 1 は、設定変更が無い場合、話者 A 変更が無いときはステップ S 1 0 3 の後の処理に移り、一方、話者 A 変更があるときは、処理を終了して（S 1 2 7）、改めて処理を実行することができる。

3. 第 3 の実施の形態

図 5 は、音声映像変換装置の第 3 の実施の形態の概略構成図である。

本実施の形態は、話者の音声言語情報を復唱者などの第 3 者が文字言語情報に変換し、それらの言語情報と話者による非言語情報とを電気通信回路を介して呈示することで、話者とユーザとの異なる言語体系間におけるコミュニケーションの補助を行うものである。

本実施の形態は、第 1 の実施の形態と同様に、特に、国際会議、多国間会議、二カ国間会議等の複数言語が関与する会議・会合・講義・授業・教育等におけるコミュニケーションを支援する。本実施の形態の音声映像変換装置は、話者用装置 1 0 0、通訳者用装置 2 0 0、第 1 及び第 2 復唱者用装置 3 0 0 及び 4 0 0、第 1 及び第 2 認識装置 5 0 0 及び 6 0 0、表示装置 7 0 0、電気通信回路 8 0 0 を備える。

話者用装置 1 0 0 は、カメラ 1 と、必要に応じてマイクを備える。通

訳者用装置 200 は、受話器及びマイクを備える。第 1 及び第 2 復唱者
用装置 300 及び 400 は、それぞれ、第 1 及び第 2 音声入力部 3 及び
4、受話器を備える。第 1 及び第 2 認識装置 500 及び 600 は、それ
ぞれ、第 1 及び第 2 音声認識部 5 及び 6、入力部 10 - b 及び 10 - c、
5 処理部 11 - b 及び 11 - c を備える。表示装置 700 は、映像遅延部
2、文字表示部 7、レイアウト設定部 8、文字映像表示部 9、入力部 1
0 - c、処理部 11 - c を備える。また、図中黒丸印●で示す構成は、
電気通信回路 800 であり、インターネット、LAN、無線 LAN、携
帯電話、PDA 等の各種電気通信回線と、電気通信回線が入力及び出力
10 される各装置 100 ~ 700 内におけるインタフェースが設けられてい
ることを表す。話者用装置 100、通訳者用装置 200、第 1 及び第 2
復唱者用装置 300 及び 400、第 1 及び第 2 認識装置 500 及び 60
0、表示装置 700 のそれぞれは、必要に応じて適宜、このような電気
通信回路 800 により接続され、音声及び／又は映像信号が通信される。
15 図中のいずれかの電気通信回路 800 を介さずに、直接有線又は無線に
より接続するようにしてもよい。よって、電気通信回線及びインターフ
ェースを有する電気通信回路 800 を用いることにより、話者 A、通訳
者 D、第 1 及び第 2 復唱者 B 及び C、第 1 及び第 2 認識装置 500 及び
600、会場などに設置される表示装置 700 は、どこに存在してもよ
く、適宜配置することができる。

カメラ 1、映像遅延部 2、第 1 及び第 2 音声入力部 3、4、第 1 音声
認識部 5、文字表示部 7、レイアウト設定部 8、文字映像表示部 9、入
力部 10 (- a、b、c) 処理部 11 (- a、b、c) の構成及び動作
は、第 1 の実施の形態の同一符号のそれと同様である。

25 ただし、入力部 10 - a は、映像遅延部 2、レイアウト設定部 8 等の
各部の各種設定、適宜のデータベースやメモリ等へのデータ入力指示を
行う。処理部 - a は、小型コンピュータであって、映像遅延部 2、入力
部 10 - a、- b 及び 10 - c、レイアウト設定部 8 等の各部を制御す
る。

また、入力部 10 - b 及び 10 - c は、第 1 及び第 2 音声認識部 5、6 の各種設定、適宜のデータベースやメモリ等へのデータ入力指示を行う。処理部 11 - b 及び 11 - c は、小型コンピュータであって、第 1 及び第 2 音声認識部 5、6 等の各部を制御する。

- 5 また、第 3 の実施の形態の音声変換処理のフローチャートは、第 1 の実施の形態と同様であり、上述したように動作する。

4. 第 4 の実施の形態

図 6 は、音声映像変換装置の第 4 の実施の形態の概略構成図である。

- 10 本実施の形態は、話者の音声言語情報を復唱者などの第 3 者が文字言語情報に変換し、それらの言語情報と話者による非言語情報とを電気通信回路を介して呈示することで、話者とユーザとの異なる言語体系間におけるコミュニケーションの補助を行うものである。

- 15 本実施の形態は、第 3 の実施の形態と同様に、特に、国際会議、多国間会議、二カ国間会議等の複数言語が関与する会議・会合・講義・授業・教育等におけるコミュニケーションを支援する。本実施の形態の音声映像変換装置は、話者用装置 100、通話者用装置 200、第 1 及び第 2 復唱者用装置 300 及び 400、第 1 認識装置 500、表示装置 700、電気通信回路 800 を備える。

20

- 第 3 の実施の形態と比較すると、第 2 音声認識部を含む第 2 認識装置 600 が省略され、第 1 認識装置 500 に選択部 20 がさらに備えられた点が、異なるが、他の構成及び動作は同様である。選択部 20 の構成及び動作は、第 2 の実施の形態と同様である。なお、第 2 音声入力部及び
25 選択部 20 は、必要に応じて、さらに省略してもよい。

また、第 4 の実施の形態の音声変換処理のフローチャートは、第 3 の実施の形態と同様であり、上述したように動作する。

5. むすび

本実施の形態では、以上のように、音声認識装置は、あらかじめ登録済みの復唱者の音声データベースを用い、話者Aの声を復唱者が復唱した音声を該音声認識装置に入力することにより音声変換し、どのような話者Aにでも高い認識率が得られるようにしている。話者Aが通訳者Dの場合でも、復唱者が通訳者Dの声を復唱することにより、外国語を高い認識率で日本語に翻訳できる。逆に、日本語で話された音声の場合は、通訳者Dが外国語に訳し、その音声をその外国語で復唱することにより、日本語を高い認識率で外国語に翻訳できる。同様に、質問者の音声も文字表示できるため双方向の会議支援を実現することができる。そのため、
5 本実施の形態は、国内会議ばかりでなく国際会議におけるコミュニケーション支援としても利用できる。

また、本実施の形態によると、話者Aの映像も取り込み、ある遅延時間で認識結果の文字列と一緒に表示する方法を採っており、話者Aの唇の動き、表情さらに手話の映像なども音声理解の手がかりとして利用することができる。聴覚障害者の読話能力に応じて、映像遅延部2による映像遅延時間を変更することができるようになっている。そのため、唇の動きを読みとる読話に習熟した聴覚障害者にとっては音声認識の5%の誤りを読話で修復できる。

20

本発明の文字映像変換方法又は文字映像変換装置・システムは、その各手順をコンピュータに実行させるための文字映像変換プログラム、文字映像変換プログラムを記録したコンピュータ読み取り可能な記録媒体、文字映像変換プログラムを含みコンピュータの内部メモリにロード可能なプログラム製品、そのプログラムを含むサーバ等のコンピュータ、等
25 により提供されることができる。

産業上の利用可能性

本発明によると、以上のように、不特定話者の声を復唱者が自己の声

に変換し音声認識装置を介して文字に変換するとともに、話者の表情などの映像を遅らせて文字とともにスクリーンなどに表示することにより、聴覚障害者などが話者の話したことを理解しやすくするための音声映像変換装置及び方法、音声映像変換プログラムを提供することができる。

5

また、本発明によると、聴覚障害者が出席するような国際会議、多国間・二国間会議等の会議において、講演者あるいは通訳者の音声を復唱者が復唱して音声認識装置に入力し、その結果である文字列を講演者の映像とともにスクリーンに表示するようにした会議支援のための音声映像変換装置及び方法、音声映像変換プログラムを提供することができる。

10

さらに、本発明によると、異種言語を使って行われる国際会議等の通訳および会議の即時印刷（情報補償）、聴覚障害者等が参加する会議や授業の支援、電話から復唱者へ声を転送し文字化された情報をユーザに提供することができる。また、本発明によると、話者とユーザとの異なる言語体系間におけるコミュニケーションの補助を行うための音声映像変換装置及び方法、音声映像変換プログラムを提供することができる。

15

また、本発明によると、さらにインターネットなどの電気通信回線を用いて通信を行う電気通信回路により、話者の声と映像を遠隔地や在宅地にいる通訳者、復唱者及び修正者に転送する手段を付加することにより、ユーザがどこにいても本システムを利用できるようにすることができる。本発明によると、介在する復唱者や通訳者が在宅ビジネスとして利用すること、さらに、外出の難しい在宅の障害者が復唱者になることにより就労を支援することができる。

20

請 求 の 範 囲

1. 話者の表情映像を撮影するカメラと、

前記カメラにより撮影された映像信号に対して予め設定された遅延時間差を与え、遅延映像データを出力する映像遅延部と、

話者が話す第1言語の内容を復唱する第1復唱者の第1言語の内容が入力される第1音声入力部と、

話者が話す第1言語の内容を通訳した通訳者の第2言語の内容をさらに復唱する第2復唱者の第2言語の内容が入力される第2音声入力部と、

10 それぞれ、前記第1及び第2音声入力部から入力された第1及び第2言語の内容を認識して第1及び第2可視言語データに変換して出力する第1及び第2音声認識部と、

前記第1及び第2音声認識部から出力された第1及び第2可視言語データと、前記映像遅延部により遅延された話者の遅延映像データとを入力し、表示状態を設定してこれらデータを同期又は略同期させた表示映像を生成するレイアウト設定部と、

前記レイアウト設定部からの出力に従い、第1及び第2可視言語データと遅延映像データとを同期又は略同期させた表示映像を表示する文字映像表示部と

20 前記第1及び第2音声認識部、前記映像遅延部、前記レイアウト設定部のいずれか又は複数の各部の各種設定を行うための入力部と、

前記第1及び第2音声認識部、前記映像遅延部、前記入力部、前記レイアウト設定部の各部を制御する処理部と
を備えた音声映像変換装置。

25

2. 話者の表情映像を撮影するカメラと、

前記カメラにより撮影された映像信号に対して予め設定された遅延時間差を与え、遅延映像データを出力する映像遅延部と、

話者又は通訳者が話す第1言語の内容を復唱する第1復唱者の第1言

語の内容が入力される第1音声入力部と、

第1音声認識部は、前記第1音声入力部から入力された第1言語の内容を認識して第1可視言語データに変換して出力する第1音声認識部と、

前記第1音声認識部から出力された第1可視言語データと、前記映像
5 遅延部により遅延された話者の遅延映像データとを入力し、表示状態を設定してこれらデータを同期又は略同期させた表示映像を生成するレイアウト設定部と、

前記レイアウト設定部からの出力に従い、第1可視言語データと遅延映像データとを同期又は略同期させた表示映像を表示する文字映像表示
10 部と

前記第1音声認識部、前記映像遅延部、前記レイアウト設定部のいずれか又は複数の各部の各種設定を行うための入力部と、

前記第1音声認識部、前記映像遅延部、前記入力部、前記レイアウト設定部の各部を制御する処理部と
15 を備えた音声映像変換装置。

3. 前記第1及び／又は第2音声認識部は、さらに、話者が話す話題又は会議の内容に従って、複数の音声認識用の言語データベースから特定の言語データベースを選択することができるようにした選択部を備えた請求項1又は2に記載の音声映像変換装置。
20

4. 前記第1及び／又は第2音声認識部は、さらに、

仮名－漢字変換において誤変換される確率を計算する誤変換確率計算部と、

25 前記誤変換確率計算部で計算された確率に応じて漢字出力か仮名文字出力かを決定する出力決定部
を備えた請求項1又は2に記載の音声映像変換装置。

5. 前記第1及び／又は第2音声認識部は、言語データベースに漢字

が登録されていない言葉を、予め定められた設定により仮名文字で表示するようにすることを特徴とする請求項 1 又は 2 に記載の音声映像変換装置。

- 5 6. 前記第 1 音声認識部により出力された第 1 言語の可視言語データを可視表示する文字表示部をさらに備えた請求項 1 又は 2 に記載の音声映像変換装置。

- 10 7. 前記レイアウト設定部は、前記文字映像表示部に表示される可視言語データ及び遅延映像データについての、単位時間当たりの行数、単位時間当たりの文字数、1 行当たりの文字数、色、大きさ、表示位置、その他の表示形式のいずれかが設定され、該設定に応じて可視言語データ及び遅延映像データについての画像処理を実行し、表示映像を生成することを特徴とする請求項 1 又は 2 に記載の音声映像変換装置。

15

8. 話者の音声を可視言語データに変換して話者の映像データとともに表示するための音声映像変換方法であって、

- 20 処理部は、入力部からの指令又は適宜の記憶部により予め定められた設定に従い、第 1 及び第 2 音声認識部及び映像遅延部の設定を行うステップと、

処理部は、入力部からの指令又は適宜の記憶部により予め定められた設定に従い、レイアウト設定部の設定を行うステップと、

カメラは、話者の映像を入力するステップと、

- 25 映像遅延部は、処理部による設定及び制御に従い、カメラに入力された映像を遅延及び必要に応じて適宜の画像処理を行い、遅延映像データを出力するステップと、

第 1 音声入力部は、話者による第 1 言語の内容を復唱する第 1 復唱者による第 1 言語の内容を入力するステップと、

第 1 音声認識部は、第 1 音声入力部に入力された第 1 復唱者による第

1 言語の内容を認識して第1可視言語データに変換するステップと、

第2音声入力部は、話者による第1言語の内容を通訳者が通訳した第2言語の内容を第2復唱者が復唱し、その復唱された第2言語の内容を入力するステップと、

5 第2音声認識部は、第2音声入力部に入力された第2復唱者による第2言語の内容を認識して第2可視言語データに変換するステップと、

レイアウト設定部は、処理部による設定及び制御に従い、第1及び第2音声認識部からの第1及び第2言語データ及び映像遅延部からの遅延映像データを入力し、それらデータの表示レイアウトを設定し、画像処理によりこれらデータを同期又は略同期させた表示映像を生成及び出力するステップと、

文字映像表示部は、レイアウト設定部からの出力に従い、第1及び第2言語データ及び映像遅延データを同期又は略同期させた表示映像を表示するステップと、

15 を含む音声映像変換方法。

9. 話者の音声を可視言語データに変換して話者の映像データとともに表示するための音声映像変換方法であって、

20 処理部は、入力部からの指令又は適宜の記憶部により予め定められた設定に従い、第1音声認識部及び映像遅延部の設定を行うステップと、

処理部は、入力部からの指令又は適宜の記憶部により予め定められた設定に従い、レイアウト設定部の設定を行うステップと、

カメラは、話者の映像を入力するステップと、

25 映像遅延部は、処理部による設定及び制御に従い、カメラに入力された映像を遅延及び必要に応じて適宜の画像処理を行い、遅延映像データを出力するステップと、

第1音声入力部は、話者又は通訳者による第1言語の内容を復唱する第1復唱者による第1言語の内容を入力するステップと、

第1音声認識部は、第1音声入力部に入力された第1復唱者による第

1 言語の内容を認識して第 1 可視言語データに変換するステップと、

レイアウト設定部は、処理部による設定及び制御に従い、第 1 音声認識部からの第 1 言語データ及び映像遅延部からの遅延映像データを入力し、それらデータの表示レイアウトを設定し、画像処理によりこれらデータ

5 ータを同期又は略同期させた表示映像を生成及び出力するステップと、

文字映像表示部は、レイアウト設定部からの出力に従い、第 1 言語データ及び映像遅延データを同期又は略同期させた表示映像を表示するステップと、

を含む音声映像変換方法。

10

10. レイアウト設定部の設定を行うステップでは、可視言語データについては、例えば、呈示文字列数、呈示文字の大きさ・フォント・色、文字列の表示位置のいずれか又は複数が設定され、また、遅延映像データについては、話者画像の大きさ、表示位置等のいずれか又は複数が設

15 定されることを特徴とする請求項 8 又は 9 に記載の音声映像変換方法。

11. さらに、文字表示部は、第 1 音声認識部から出力された第 1 可視言語データを表示するステップとを含むことを特徴とする請求項 8 又は 9 に記載の音声映像変換方法。

20

12. 話者の音声を可視言語データに変換して話者の映像データとともに表示するための音声映像変換プログラムであって、

処理部は、入力部からの指令又は適宜の記憶部により予め定められた設定に従い、第 1 及び第 2 音声認識部及び映像遅延部の設定を行うステ

25 ップと、

処理部は、入力部からの指令又は適宜の記憶部により予め定められた設定に従い、レイアウト設定部の設定を行うステップと、

カメラは、話者の映像を入力するステップと、

映像遅延部は、処理部による設定及び制御に従い、カメラに入力され

た映像を遅延及び必要に応じて適宜の画像処理を行い、遅延映像データ
を出力するステップと、

第1音声入力部は、話者による第1言語の内容を復唱する第1復唱者
による第1言語の内容を入力するステップと、

5 第1音声認識部は、第1音声入力部に入力された第1復唱者による第
1言語の内容を認識して第1可視言語データに変換するステップと、

第2音声入力部は、話者による第1言語の内容を通訳者が通訳した第
2言語の内容を第2復唱者が復唱し、その復唱された第2言語の内容を
入力するステップと、

10 第2音声認識部は、第2音声入力部に入力された第2復唱者による第
2言語の内容を認識して第2可視言語データに変換するステップと、

レイアウト設定部は、処理部による設定及び制御に従い、第1及び第
2音声認識部からの第1及び第2言語データ及び映像遅延部からの遅延
映像データを入力し、それらデータの表示レイアウトを設定し、画像処
15 理によりこれらデータを同期又は略同期させた表示映像を生成及び出力
するステップと、

文字映像表示部は、レイアウト設定部からの出力に従い、第1及び第
2言語データ及び映像遅延データを同期又は略同期させた表示映像を表
示するステップと、

20 をコンピュータに実行させるための音声映像変換方法プログラム。

13. 話者の音声を可視言語データに変換して話者の映像データとと
もに表示するための音声映像変換プログラムであって、

25 処理部は、入力部からの指令又は適宜の記憶部により予め定められた
設定に従い、第1音声認識部及び映像遅延部の設定を行うステップと、

処理部は、入力部からの指令又は適宜の記憶部により予め定められた
設定に従い、レイアウト設定部の設定を行うステップと、

カメラは、話者の映像を入力するステップと、

映像遅延部は、処理部による設定及び制御に従い、カメラに入力され

た映像を遅延及び必要に応じて適宜の画像処理を行い、遅延映像データを出力するステップと、

第1音声入力部は、話者又は通訳者による第1言語の内容を復唱する第1復唱者による第1言語の内容を入力するステップと、

5 第1音声認識部は、第1音声入力部に入力された第1復唱者による第1言語の内容を認識して第1可視言語データに変換するステップと、

レイアウト設定部は、処理部による設定及び制御に従い、第1音声認識部からの第1言語データ及び映像遅延部からの遅延映像データを入力し、それらデータの表示レイアウトを設定し、画像処理によりこれらデータを同期又は略同期させた表示映像を生成及び出力するステップと、

10 文字映像表示部は、レイアウト設定部からの出力に従い、第1言語データ及び映像遅延データを同期又は略同期させた表示映像を表示するステップと、

をコンピュータに実行させるための音声映像変換方法プログラム。

15

14. 話者が話す第1言語の内容を復唱する第1復唱者の第1言語の内容を認識して第1可視言語データに変換して出力する第1音声認識部と、前記第1音声認識部の各種設定を行うための第1入力部と、前記第1音声認識部及び前記第1入力部を制御する第1処理部とを有する第1認識装置と、

20

話者が話す第1言語の内容を通訳した通訳者の第2言語の内容をさらに復唱する第2復唱者の第2言語の内容を認識して第2可視言語データに変換して出力する第2音声認識部と、前記第2音声認識部の各種設定を行うための第2入力部と、前記第2音声認識部及び前記第2入力部を

25

制御する第2処理部とを有する第2認識装置と、

前記第1及び第2認識装置からの出力が入力され、文字及び映像を表示するための表示装置と

を備え、

前記表示装置は、

カメラにより撮影された映像信号に対して予め設定された遅延時間差を与え、遅延映像データを出力する映像遅延部と、

前記第1及び第2認識装置から出力された第1及び第2可視言語データと、前記映像遅延部により遅延された話者の遅延映像データとを入力し、表示状態を設定してこれらデータを同期又は略同期させた表示映像を生成するレイアウト設定部と、

前記レイアウト設定部から出力された表示映像を表示する文字映像表示部と

前記映像遅延部及び前記レイアウト設定部の各種設定を行うための第3入力部と、

前記映像遅延部、前記第3入力部、前記レイアウト設定部の各部を制御する第3処理部と
を有する音声映像変換装置。

15 15. 話者又は通訳者が話す第1言語の内容を復唱する第1復唱者の第1言語の内容を認識して第1可視言語データに変換して出力する第1音声認識部と、前記第1音声認識部の各種設定を行うための第1入力部と、前記第1音声認識部及び前記第1入力部を制御する第1処理部とを有する第1認識装置と、

20 前記第1認識装置からの出力が入力され、文字及び映像を表示するための表示装置と
を備え、

前記表示装置は、

カメラにより撮影された映像信号に対して予め設定された遅延時間差を与え、遅延映像データを出力する映像遅延部と、

25 前記第1認識装置から出力された第1可視言語データと、前記映像遅延部により遅延された話者の遅延映像データとを入力し、表示状態を設定してこれらデータを同期又は略同期させた表示映像を生成するレイアウト設定部と、

前記レイアウト設定部から出力された表示映像を表示する文字映像表示部と

前記映像遅延部及び前記レイアウト設定部の各種設定を行うための第3入力部と、

- 5 前記映像遅延部、前記第3入力部、前記レイアウト設定部の各部を制御する第3処理部と
を有する音声映像変換装置。

16. 話者の表情映像を撮影するカメラと、

- 10 話者の音声を入力する入力部と、

電気通信回線を介して通信するためのインタフェースと
を有し、電気通信回線及びインタフェースを経て音声及び映像信号を出力する話者用装置をさらに備えた請求項14又は15に記載の音声映像変換装置。

15

17. 話者が話す第1言語の内容を復唱する第1復唱者の第1言語の内容が入力される第1音声入力部と、電気通信回線を介して通信するためのインタフェースとを有し、電気通信回線及びインタフェースを経て音声信号を前記第1認識装置に出力する第1復唱者用装置をさらに備えた請求項14又は15に記載の音声映像変換装置。

20

18. 話者が話す第1言語の内容を通訳した通訳者の第2言語の内容をさらに復唱する第2復唱者の第2言語の内容が入力される第2音声入力部と、電気通信回線を介して通信するためのインタフェースとを有し、
25 電気通信回線及びインタフェースを経て音声信号を前記第2認識装置に出力する第2復唱者用装置さらに備えた請求項14又は15に記載の音声映像変換装置。

19. 前記第1及び第2認識装置及び前記表示装置は、電気通信回線

を介して通信するためのインタフェースと、

前記第 1 及び／又は第 2 認識装置の出力は、電気通信回線及びインタフェースを介して前記表示装置に伝送されることを特徴とする請求項 14 又は 15 に記載の音声映像変換装置。

5

20. 前記レイアウト設定部は、前記文字映像表示部に表示される可視言語データ及び遅延映像データについての、単位時間当たりの行数、単位時間当たりの文字数、1 行当たりの文字数、色、大きさ、表示位置、その他の表示形式のいずれかが設定され、該設定に応じて可視言語データ及び遅延映像データについての画像処理を実行し、表示映像を生成することを特徴とする請求項 14 又は 15 に記載の音声映像変換装置。

10

21. 話者の音声を可視言語データに変換して話者の映像データとともに表示するための音声映像変換方法であって、

15

第 1 及び第 2 処理部、第 3 処理部は、それぞれ、第 1 及び第 2 入力部、第 3 入力部からの指令又は適宜の記憶部により予め定められた設定に従い、第 1 及び第 2 音声認識部、映像遅延部の設定を行うステップと、

第 3 処理部は、第 3 入力部からの指令又は適宜の記憶部により予め定められた設定に従い、レイアウト設定部の設定を行うステップと、

20

映像遅延部は、第 3 処理部による設定及び制御に従い、カメラに入力された話者の映像を遅延及び必要に応じて適宜の画像処理を行い、遅延映像データを出力するステップと、

第 1 音声認識部は、話者による第 1 言語の内容を復唱する第 1 復唱者による第 1 言語の内容を認識して第 1 可視言語データに変換するステップと、

25

第 2 音声認識部は、話者による第 1 言語の内容を通訳者が通訳した第 2 言語の内容を復唱した第 2 復唱者による第 2 言語の内容を認識して第 2 可視言語データに変換するステップと、

レイアウト設定部は、第 3 処理部による設定及び制御に従い、第 1 及

び第 2 音声認識部からの第 1 及び第 2 可視言語データ及び映像遅延部からの遅延映像データを入力し、それらデータの表示レイアウトを設定し、画像処理によりこれらデータを同期又は略同期させた表示映像を生成及び出力するステップと、

- 5 文字映像表示部は、レイアウト設定部からの出力に従い、第 1 及び第 2 可視言語データ及び映像遅延データを同期又は略同期させた表示映像を表示するステップと、
を含む音声映像変換方法。

- 10 2 2. 話者の音声を可視言語データに変換して話者の映像データとともに表示するための音声映像変換方法であって、

第 1 及び第 3 処理部は、それぞれ、第 1 及び第 3 入力部からの指令又は適宜の記憶部により予め定められた設定に従い、第 1 音声認識部及び映像遅延部の設定を行うステップと、

- 15 第 3 処理部は、第 3 入力部からの指令又は適宜の記憶部により予め定められた設定に従い、レイアウト設定部の設定を行うステップと、

映像遅延部は、第 3 処理部による設定及び制御に従い、カメラに入力された話者の映像を遅延及び必要に応じて適宜の画像処理を行い、遅延映像データを出力するステップと、

- 20 第 1 音声認識部は、話者又は通訳者による第 1 言語の内容を復唱する第 1 復唱者による第 1 言語の内容を認識して第 1 可視言語データに変換するステップと、

- レイアウト設定部は、第 3 処理部による設定及び制御に従い、第 1 音声認識部からの第 1 可視言語データ及び映像遅延部からの遅延映像データを入力し、それらデータの表示レイアウトを設定し、画像処理によりこれらデータを同期又は略同期させた表示映像を生成及び出力するステップと、
- 25

文字映像表示部は、レイアウト設定部からの出力に従い、第 1 可視言語データ及び映像遅延データを同期又は略同期させた表示映像を表示す

るステップと、
を含む音声映像変換方法。

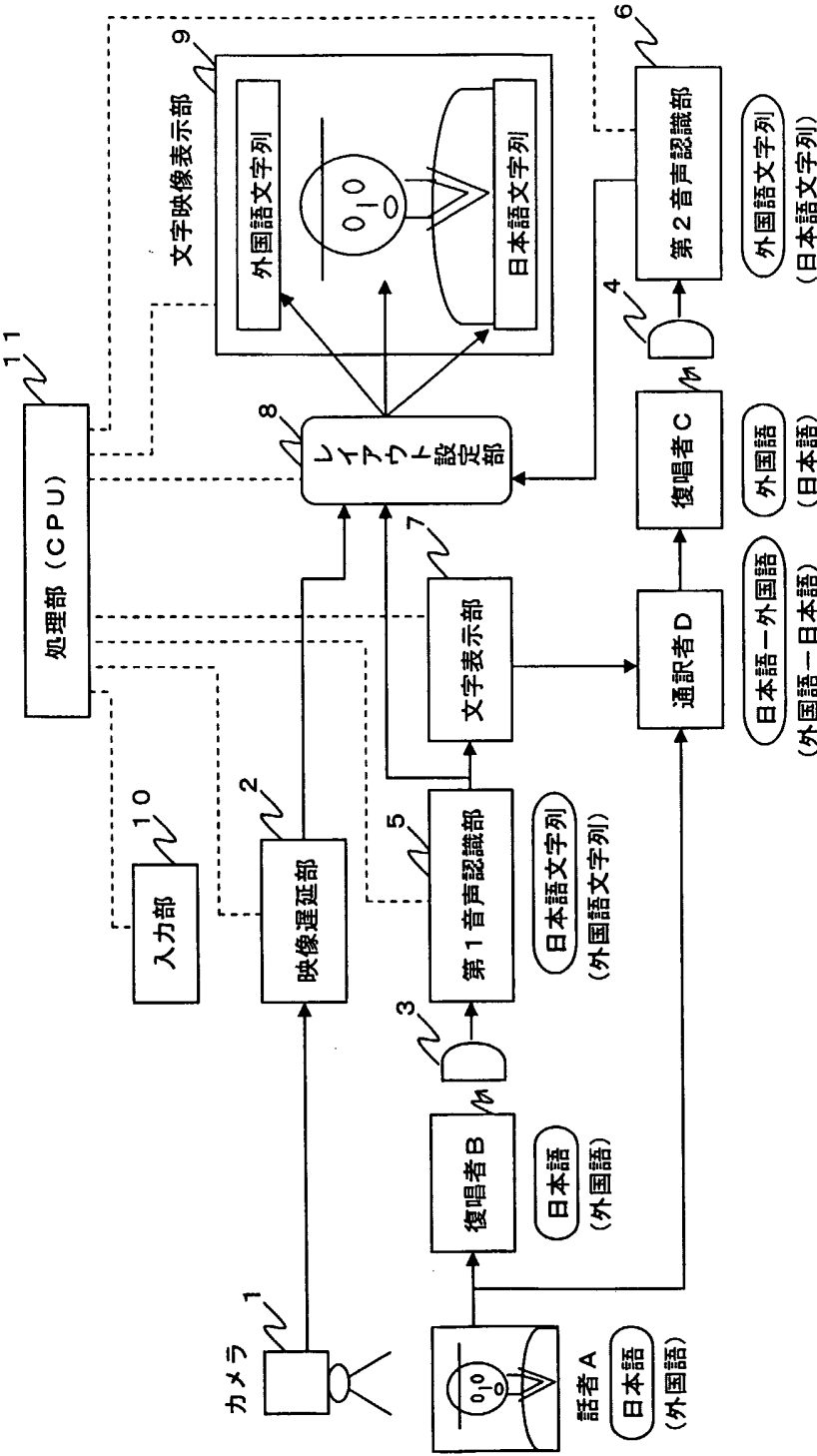
23. レイアウト設定部の設定を行うステップでは、可視言語データ
5 については、例えば、呈示文字列数、呈示文字の大きさ・フォント・色、
文字列の表示位置のいずれか又は複数が設定され、また、遅延映像データ
については、話者画像の大きさ、表示位置等のいずれか又は複数が設定
されることを特徴とする請求項8又は9に記載の音声映像変換方法。

10 24. 話者による第1言語の内容及びカメラから入力された話者の映像を、
電気通信回路を介して伝送するステップをさらに含む請求項8又は9
に記載の音声映像変換方法。

25. 第1復唱者による第1言語の内容、第2復唱者による第2言語
15 の内容、通訳者による第2言語の内容、のいずれか又は複数を電気通信
回路を介して伝送するステップをさらに含む請求項8又は9に記載の音声
映像変換方法。

26. 第1及び／又は第2音声認識装置から出力された第1及び／又は
20 は第2可視言語データを電気通信回路を介して入力するステップをさらに
含む請求項8又は9に記載の音声映像変換方法。

図 1



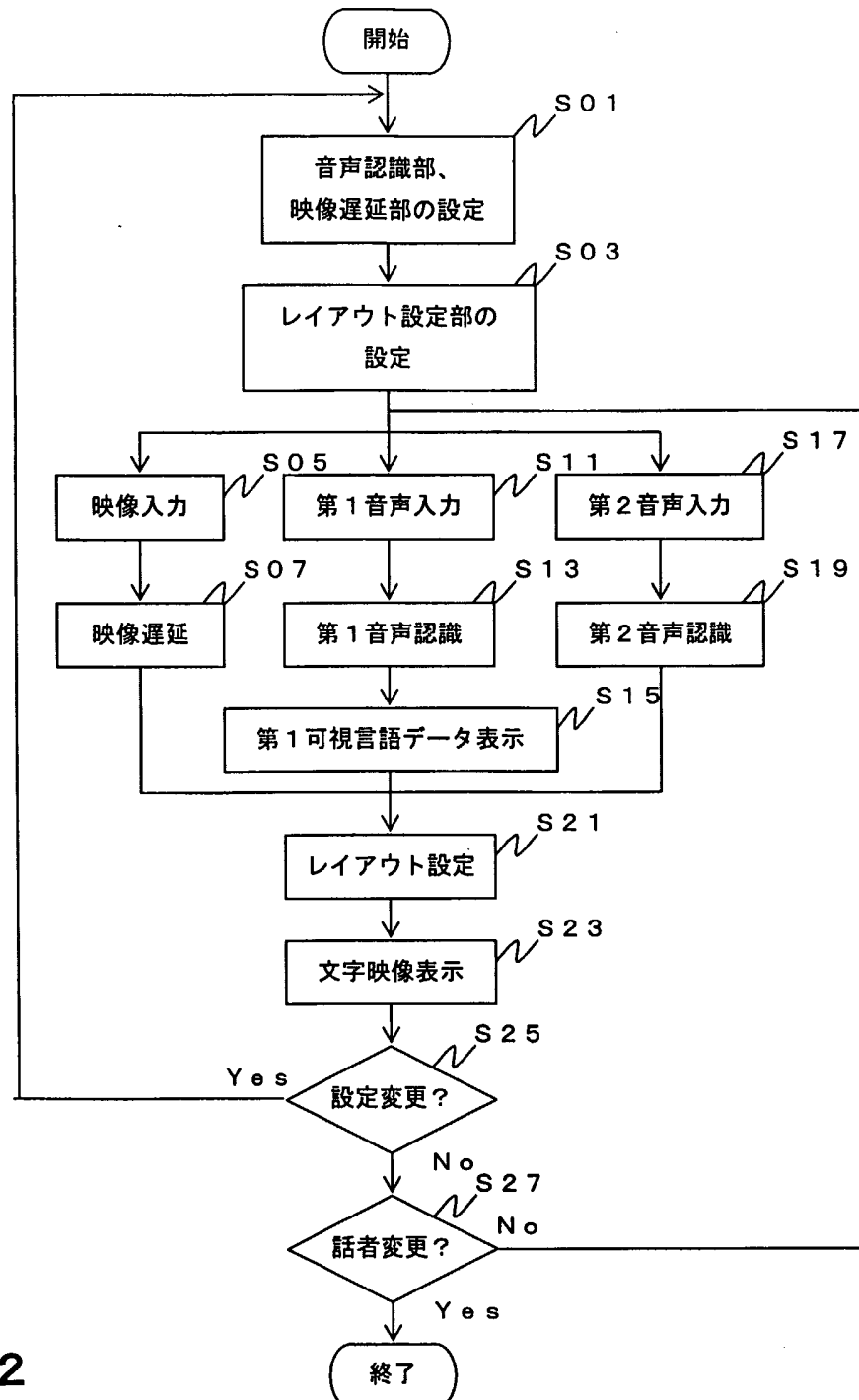
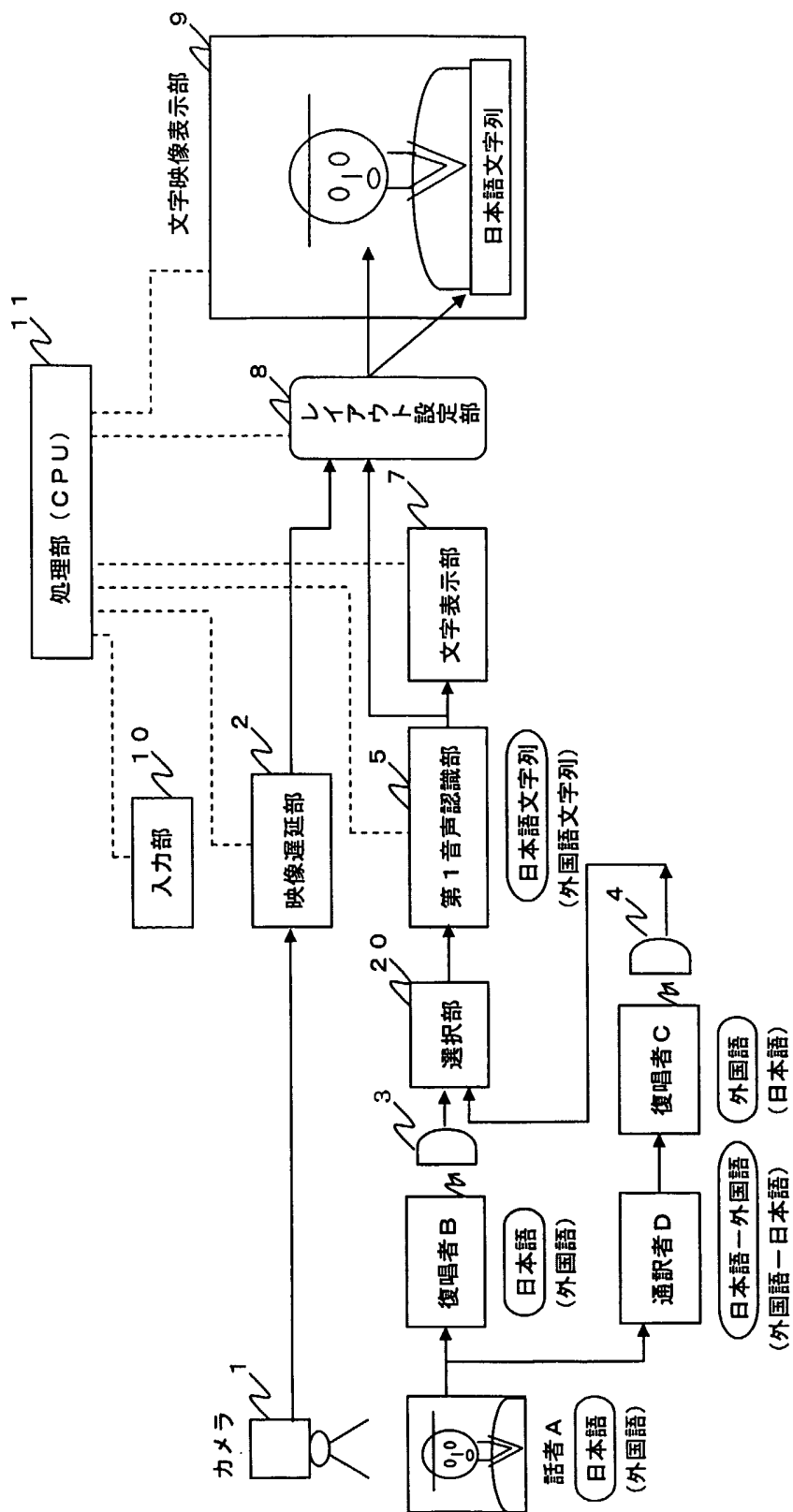


図 2

图 3



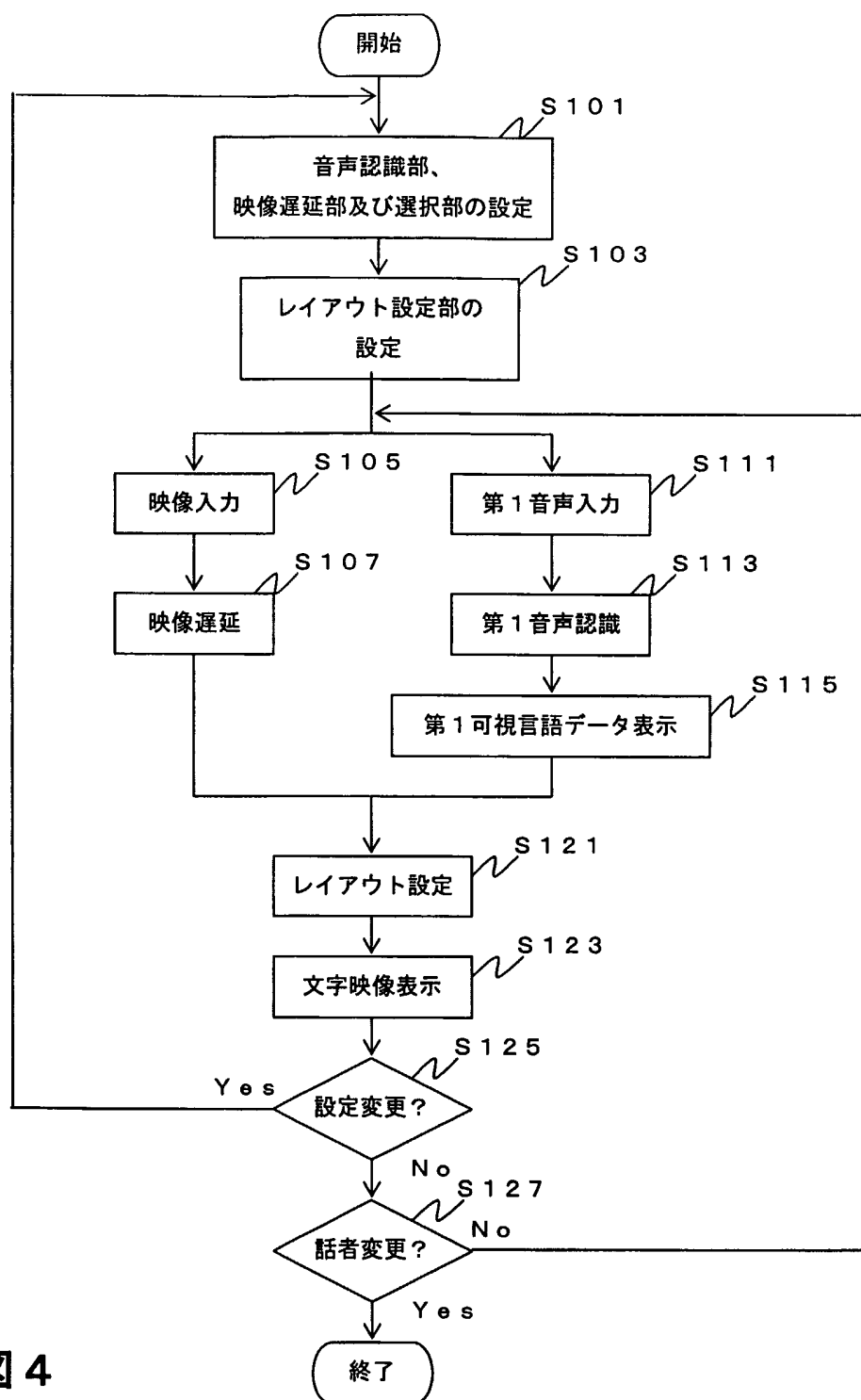


図 4

図 5

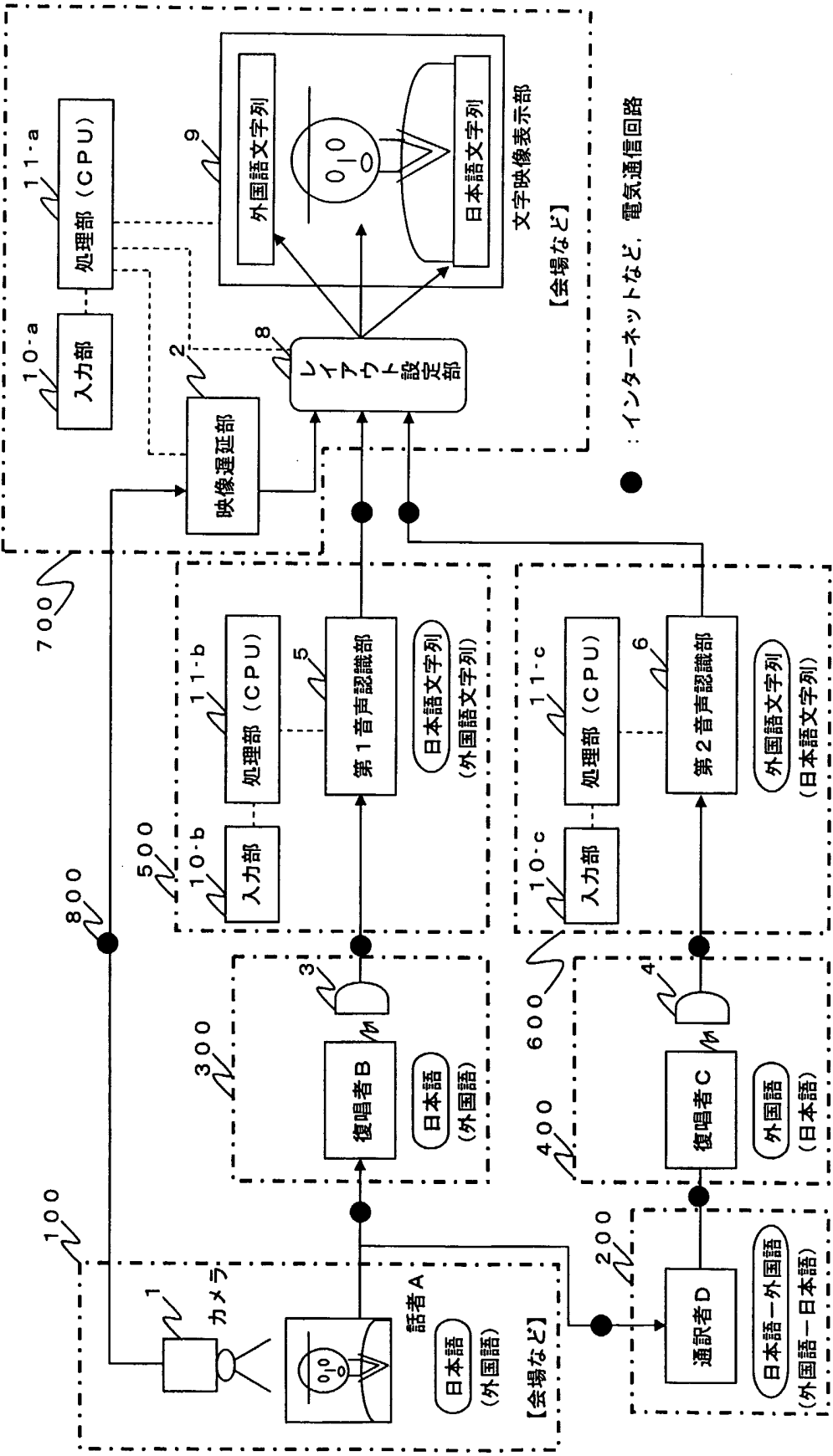
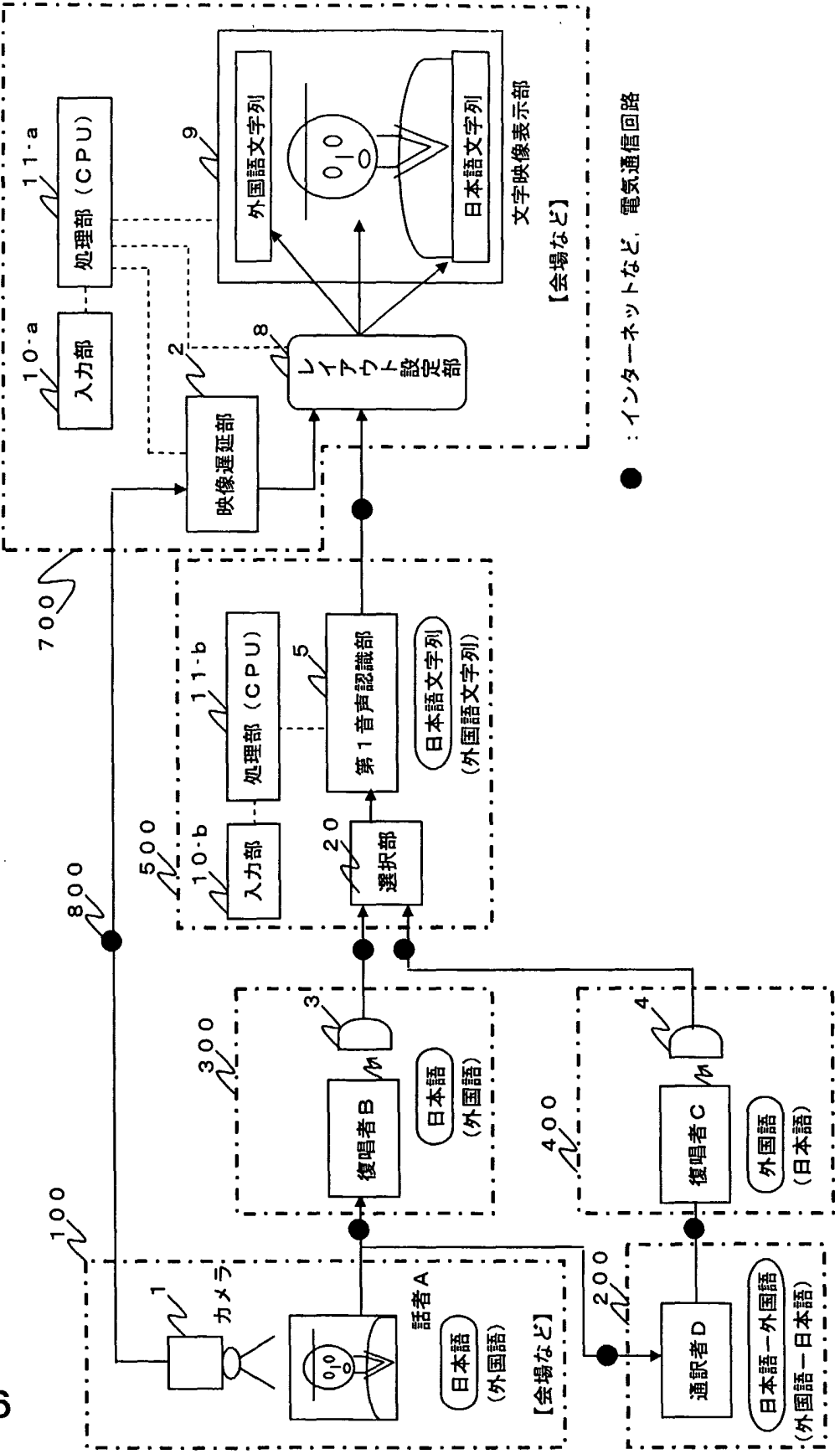


図 6



特許協力条約に基づく国際出願願書

原本（出願用） - 印刷日時 2003年03月19日（19. 03. 2003）水曜日 11時39分32秒

VIII-5-1	不利にならない開示又は新規性喪失の例外に関する申立て 不利にならない開示又は新規性喪失の例外に関する申立て（規則4.17(v)及び51の2.1(a)(v)）	本国際出願に関し、 科学技術振興事業団は、 本国際出願の請求項に記載された対象が以下のよう に開示されたことを申し立てる。
VIII-5-1 (i) VIII-5-1 (ii) VIII-5-1 (iii) VIII-5-1 (iv)	開示の種類 開示の日付： 開示の名称： 開示の場所：	刊行物 2001年10月26日（26. 10. 2001） 北海道新聞
VIII-5-1 (i) VIII-5-1 (ii) VIII-5-1 (iii) VIII-5-1 (iv)	開示の種類 開示の日付： 開示の名称： 開示の場所：	刊行物 2001年11月02日（02. 11. 2001） 北海道新聞
VIII-5-1 (i) VIII-5-1 (ii) VIII-5-1 (iii) VIII-5-1 (iv)	開示の種類 開示の日付： 開示の名称： 開示の場所：	刊行物 2001年11月15日（15. 11. 2001） 日刊工業新聞
VIII-5-1 (i) VIII-5-1 (ii) VIII-5-1 (iii) VIII-5-1 (iv)	開示の種類 開示の日付： 開示の名称： 開示の場所：	刊行物 2002年01月16日（16. 01. 2002） 産経新聞
VIII-5-1 (i) VIII-5-1 (ii) VIII-5-1 (iii) VIII-5-1 (iv)	開示の種類 開示の日付： 開示の名称： 開示の場所：	刊行物 2002年01月16日（16. 01. 2002） 熊本日日新聞
VIII-5-1 (i) VIII-5-1 (ii) VIII-5-1 (iii) VIII-5-1 (iv)	開示の種類 開示の日付： 開示の名称： 開示の場所：	刊行物 2002年01月16日（16. 01. 2002） 埼玉新聞
VIII-5-1 (i) VIII-5-1 (ii) VIII-5-1 (iii)	開示の種類 開示の日付： 開示の名称：	刊行物 2002年01月16日（16. 01. 2002） 室蘭民報

VIII-5-1 (iv)	開示の場所:	
VIII-5-1 (i)	開示の種類	刊行物
VIII-5-1 (ii)	開示の日付:	2002年01月18日 (18. 01. 2002)
VIII-5-1 (iii)	開示の名称:	秋田魁新報
VIII-5-1 (iv)	開示の場所:	
VIII-5-1 (i)	開示の種類	刊行物
VIII-5-1 (ii)	開示の日付:	2002年01月18日 (18. 01. 2002)
VIII-5-1 (iii)	開示の名称:	Japan Times
VIII-5-1 (iv)	開示の場所:	

特許協力条約に基づく国際出願願書

原本（出願用） - 印刷日時 2003年03月19日（19. 03. 2003）水曜日 11時39分32秒

VIII-5-1 (i)	開示の種類	刊行物
VIII-5-1 (ii)	開示の日付:	2002年12月02日 (02. 12. 2002)
VIII-5-1 (iii)	開示の名称:	日本経済新聞
VIII-5-1 (iv)	開示の場所:	
VIII-5-1 (i)	開示の種類	刊行物
VIII-5-1 (ii)	開示の日付:	2002年11月30日 (30. 11. 2002)
VIII-5-1 (iii)	開示の名称:	札幌タイムス
VIII-5-1 (iv)	開示の場所:	
VIII-5-1 (v)	本申立ては、次の指定国のため になされたものである。:	すべての指定国

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP03/03305

A. CLASSIFICATION OF SUBJECT MATTER

Int.Cl⁷ G10L15/00, H04N5/278, G06F17/28

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Int.Cl⁷ G10L15/00-15/28, H04N5/278, G06F17/28

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho	1926-1995	Toroku Jitsuyo Shinan Koho	1994-2003
Kokai Jitsuyo Shinan Koho	1971-2003	Jitsuyo Shinan Toroku Koho	1996-2003

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

JICST FILE (JOIS)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	KOBAYASHI et al., "Chokaku Shogaisha no tame no Onsei Ninshiki o Katsuyo shita Real Time Jimaku Sonyu System (2)", The Institute of Electronics, Information and Communication Engineers Gijutsu Kenkyu Hokoku, 20 January, 2001 (20.01.01), Vol.100, No.600, ET2000-108, pages 129 to 134	2-3, 5-7, 9-11, 13, 15-17, 20, 22-24
Y	JP 2002-10138 A (Nippon Telegraph And Telephone Corp.), 11 January, 2002 (11.01.02), Par. No. [0003] (Family: none)	2-3, 5-7, 9-11, 13, 15-17, 20, 22-24
Y	JP 10-234016 A (Hitachi, Ltd.), 02 September, 1998 (02.09.98), Claim 5 (Family: none)	2-3, 5-7, 9-11, 13, 15-17, 20, 22-24

☒ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:	"I" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier document but published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search
25 April, 2003 (25.04.03)

Date of mailing of the international search report
20 May, 2003 (20.05.03)

Name and mailing address of the ISA/
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP03/03305

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	JP 63-219067 A (Director General, Agency of Industrial Science and Technology), 12 September, 1988 (12.09.88), Full text; all drawings (Family: none)	3
Y	JP 8-50698 A (Mazda Motor Corp.), 20 February, 1996 (20.02.96), Par. No. [0017]; Fig. 8 (Family: none)	3
T	KATO et al., "Onsei Ninshiki Gijutsu o Mochiita Chokaku Shogaisha Muke no Kokusai Kaigi Sanka Shien System no Sekkei", The Japan Society of Mechanical Engineers [No.02-6] Robotics·Mechatronics Koenkai '02 Koen Ronbubshu, 07 June, 2002 (07.06.02), 1A1-C08, pages 1A1-C08 (1)-1A1-C08(2)	1-26

A. 発明の属する分野の分類 (国際特許分類 (IPC))

Int. Cl⁷ G10L15/00, H04N5/278, G06F17/28

B. 調査を行った分野

調査を行った最小限資料 (国際特許分類 (IPC))

Int. Cl⁷ G10L15/00-15/28, H04N5/278, G06F17/28

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報 1926~1995年

日本国公開実用新案公報 1971~2003年

日本国登録実用新案公報 1994~2003年

日本国実用新案登録公報 1996~2003年

国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)

JICSTファイル (JOIS)

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
Y	小林他, 「聴覚障害者のための音声認識を活用したリアルタイム字幕挿入システム (2)」, 電子情報通信学会技術研究報告, 2001.01.20, Vol.100, No.600, ET2000-108, Pages 129 - 134	2-3, 5-7, 9-11, 13, 15-17, 20, 22-24
Y	JP 2002-10138 A (日本電信電話株式会社) 2002.01.11, 【0003】 (ファミリーなし)	2-3, 5-7, 9-11, 13, 15-17, 20, 22-24

☒ C欄の続きにも文献が列挙されている。☐ パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー

「A」 特に関連のある文献ではなく、一般的技術水準を示すもの

「E」 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの

「L」 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)

「O」 口頭による開示、使用、展示等に言及する文献

「P」 国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献

「T」 国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの

「X」 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの

「Y」 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの

「&」 同一パテントファミリー文献

国際調査を完了した日

25.04.03

国際調査報告の発送日

20.05.03

国際調査機関の名称及びあて先

日本国特許庁 (ISA/JP)

郵便番号100-8915

東京都千代田区霞が関三丁目4番3号

特許庁審査官 (権限のある職員)

榎本 剛



5C

9379

電話番号 03-3581-1101 内線 3541

C (続き) . 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
Y	J P 10-234016 A (株式会社日立製作所) 1998.09.02, 請求項5 (ファミリーなし)	2-3, 5-7, 9-11, 13, 15-17, 20, 22-24
Y	J P 63-219067 A (工業技術院長) 1988.09.12, 全文, 全図 (ファミリーなし)	3
Y	J P 8-50698 A (マツダ株式会社) 1996.02.20, 【0017】, 第8図 (ファミリーなし)	3
T	加藤他, 「音声認識技術を用いた聴覚障害者向けの国際会議参加支援システムの設計」, 日本機械学会 [No. 02-6] ロボティクス・メカトロニクス講演会 '02 講演論文集, 2002.06.07, 1A1-C08, Pages 1A1-C08(1) - 1A1-C08(2)	1-26