

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局



(10) 国際公開番号

WO 2011/040056 A1

(43) 国際公開日

2011年4月7日(07.04.2011)

PCT

(51) 国際特許分類:

G10L 15/10 (2006.01) G10L 15/00 (2006.01)
G06F 3/16 (2006.01) G10L 15/22 (2006.01)
G06F 17/28 (2006.01) G10L 15/28 (2006.01)
G10L 13/00 (2006.01)

(21) 国際出願番号: PCT/JP2010/053419

(22) 国際出願日: 2010年3月3日(03.03.2010)

(25) 国際出願の言語: 日本語

(26) 国際公開の言語: 日本語

(30) 優先権データ:
特願 2009-230442 2009年10月2日(02.10.2009) JP

(71) 出願人(米国を除く全ての指定国について): 独立行政法人情報通信研究機構(National Institute of Information and Communications Technology) [JP/JP]; 〒1848795 東京都小金井市貫井北町4-2-1 Tokyo (JP).

(72) 発明者: および

(75) 発明者/出願人(米国についてのみ): 中村 哲 (NAKAMURA, Satoshi) [JP/JP]; 〒1848795 東京都小金井市貫井北町4-2-1 独立行政法人情報

報通信研究機構内 Tokyo (JP). 隅田 英一郎 (SUMITA, Eiichiro) [JP/JP]; 〒1848795 東京都小金井市貫井北町4-2-1 独立行政法人情報通信研究機構内 Tokyo (JP). 葦 莉 豊 (ASHIKARI, Yutaka) [JP/JP]; 〒1848795 東京都小金井市貫井北町4-2-1 独立行政法人情報通信研究機構内 Tokyo (JP). 木村 法幸 (KIMURA, Noriyuki) [JP/JP]; 〒1848795 東京都小金井市貫井北町4-2-1 独立行政法人情報通信研究機構内 Tokyo (JP). 堀 智織 (HORI, Chiori) [JP/JP]; 〒1848795 東京都小金井市貫井北町4-2-1 独立行政法人情報通信研究機構内 Tokyo (JP).

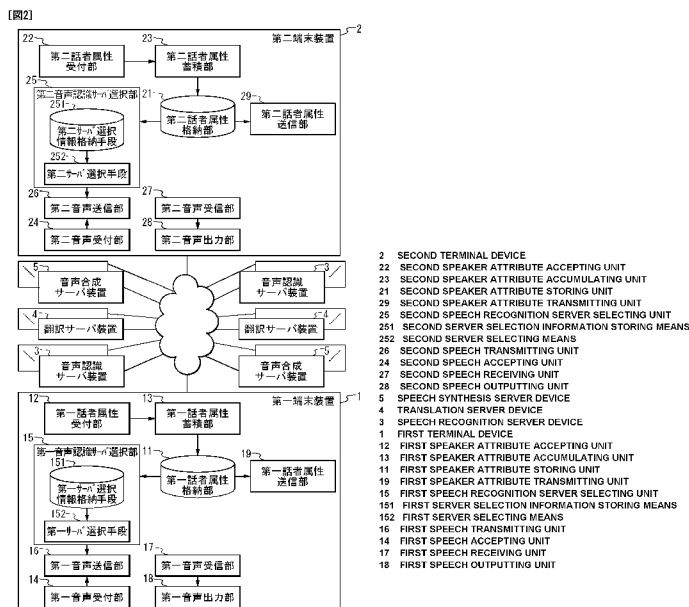
(74) 代理人: 谷川 英和(TANIGAWA, Hidekazu); 〒5400008 大阪府大阪市中央区大手前1丁目7-31 OMMビル8階 私書箱53号 Osaka (JP).

(81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR,

[続葉有]

(54) Title: SPEECH TRANSLATION SYSTEM, FIRST TERMINAL DEVICE, SPEECH RECOGNITION SERVER DEVICE, TRANSLATION SERVER DEVICE, AND SPEECH SYNTHESIS SERVER DEVICE

(54) 発明の名称: 音声翻訳システム、第一端末装置、音声認識サーバ装置、翻訳サーバ装置、および音声合成サーバ装置



(57) Abstract: In a conventional network-type speech translation system, devices and models which perform speech recognition, translation, and speech synthesis cannot be changed on the basis of the attributes of a speaker, causing deterioration in precision or inappropriate output in each process of the speech recognition, translation, and speech synthesis. Disclosed is a network-type speech translation system which on the basis of the attributes of the speaker, appropriately changes a speech recognition server device and a translation server device, and appropriately changes a speech synthesis server device and a speech synthesis model, thereby enabling improvement in precision and appropriate output in each process of the speech recognition, translation, and speech synthesis.

(57) 要約: 【課題】従来のネットワーク型の音声翻訳システムにおいて、話者の属性に応じて音声認識や翻訳や音声合成を行う装置やモデルを変更できなかったために、音声認識や翻訳や音声合成の各処理において、精度が低下したり、不適切な出力が生じたりしていた。【解決手段】ネットワーク型の音声翻訳システムにおいて、話者の属性に基づいて、音声認識を行うサーバ装置や音声認識モデルを適切に変更したり、音声合成を行うサーバ装置や音声合成モデルを適切に変更したりすることにより、音声認識、または翻訳、または音声合成の各処理において、精度が向上し、適切な出力が行われる。



LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

- (84) 指定国 (表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ

添付公開書類:

— 国際調査報告 (条約第 21 条(3))

明 細 書

発明の名称：

音声翻訳システム、第一端末装置、音声認識サーバ装置、翻訳サーバ装置、および音声合成サーバ装置

技術分野

[0001] 本発明は、音声翻訳を行う音声翻訳システム等に関するものである。

背景技術

[0002] 従来の音声翻訳システムにおいて、音声認識の精度を向上させたり、翻訳処理の精度を向上させたり等、各部分処理の精度を向上させるための技術が存在した（例えば、特許文献1、特許文献2参照）。

特許文献1：特開2008-243080号公報（第1頁、第1図等）

特許文献2：特開2009-140503号公報（第1頁、第1図等）

発明の開示

発明が解決しようとする課題

[0003] しかしながら、従来の音声翻訳システムにおいては、音声翻訳において、話者の属性に応じて音声認識や翻訳や音声合成を行う装置やモデルを変更できなかったために、音声認識や翻訳や音声合成の各処理において、精度が低下したり、不適切な出力が生じたりしていた。

課題を解決するための手段

[0004] 本第一の発明の音声翻訳システムは、音声を入力する第一端末装置、2以上の音声認識サーバ装置、1以上の翻訳サーバ装置、1以上の音声合成サーバ装置とを有する音声翻訳システムであって、第一端末装置は、話者の属性値である話者属性を、1以上格納し得る第一話者属性格納部と、音声を受け付ける第一音声受付部と、1以上の話者属性に応じて、2以上の音声認識サーバ装置のうちの音声認識サーバ装置を選択する第一音声認識サーバ選択部と、第一音声認識サーバ選択部が選択した音声認識サーバ装置に、第一音声受付部が受け付けた音声から構成される音声情報を送信する第一音声送信

部とを具備し、音声認識サーバ装置は、2以上の言語のうちのすべての言語または2以上の一部の言語について、音声認識モデルを格納し得る音声認識モデル格納部と、音声情報を受信する音声情報受信部と、音声情報受信部が受信した音声情報を、音声認識モデル格納部の音声認識モデルを用いて音声認識し、音声認識結果を取得する音声認識部と、音声認識結果を送信する音声認識結果送信部とを具備し、翻訳サーバ装置は、2以上の言語のうちのすべての言語または2以上の一部の言語について、翻訳モデルを格納し得る翻訳モデル格納部と、音声認識結果を受信する音声認識結果受信部と、音声認識結果受信部が受信した音声認識結果を、翻訳モデル格納部の翻訳モデルを用いて、目的言語に翻訳し、翻訳結果を取得する翻訳部と、翻訳結果を送信する翻訳結果送信部とを具備し、音声合成サーバ装置は、2以上の言語のうちのすべての言語または2以上の一部の言語について、音声合成モデルを格納し得る音声合成モデル格納部と、翻訳結果を受信する翻訳結果受信部と、翻訳結果受信部が受信した翻訳結果を、音声合成モデル格納部の音声合成モデルを用いて音声合成し、音声合成結果を取得する音声合成部と、音声合成結果を第二端末装置に送信する音声合成結果送信部とを具備する音声翻訳システムである。

[0005] かかる構成により、音声翻訳において、話者の属性に応じて音声認識を行う装置を変更でき、音声認識処理において、精度が向上する。

[0006] また、本第二の発明の音声翻訳システムは、音声を入力する第一端末装置、1以上の音声認識サーバ装置、1以上の翻訳サーバ装置、1以上の音声合成サーバ装置とを有する音声翻訳システムであって、第一端末装置は、音声を受け付ける第一音声受付部と、第一音声受付部が受け付けた音声から構成される音声情報を、音声認識サーバ装置に送信する第一音声送信部とを具備し、音声認識サーバ装置は、話者の属性値である話者属性を、1以上格納し得る第三話者属性格納部と、2以上の言語のうちのすべての言語または2以上の一部の言語について、2以上の音声認識モデルを格納し得る音声認識モデル格納部と、音声情報を受信する音声情報受信部と、1以上の話者属性に

応じて、2以上の音声認識モデルから、一の音声認識モデルを選択する音声認識モデル選択部と、音声情報受信部が受信した音声情報を、音声認識モデル選択部が選択した音声認識モデルを用いて音声認識し、音声認識結果を取得する音声認識部と、音声認識結果を送信する音声認識結果送信部とを具備し、翻訳サーバ装置は、2以上の言語のうちのすべての言語または2以上の一部の言語について、翻訳モデルを格納し得る翻訳モデル格納部と、音声認識結果を受信する音声認識結果受信部と、音声認識結果受信部が受信した音声認識結果を、翻訳モデル格納部の翻訳モデルを用いて、目的言語に翻訳し、翻訳結果を取得する翻訳部と、翻訳結果を送信する翻訳結果送信部とを具備し、音声合成サーバ装置は、2以上の言語のうちのすべての言語または2以上の一部の言語について、音声合成モデルを格納し得る音声合成モデル格納部と、翻訳結果を受信する翻訳結果受信部と、翻訳結果受信部が受信した翻訳結果を、音声合成モデル格納部の音声合成モデルを用いて音声合成し、音声合成結果を取得する音声合成部と、音声合成結果を第二端末装置に送信する音声合成結果送信部とを具備する音声翻訳システムである。

[0007] かかる構成により、音声翻訳において、話者の属性に応じて音声認識を行うモデルを変更でき、音声認識処理において、精度が向上する。

[0008] また、本第三の発明の音声翻訳システムは、1以上の音声認識サーバ装置、2以上の翻訳サーバ装置、1以上の音声合成サーバ装置とを有する音声翻訳システムであって、音声認識サーバ装置は、話者の属性値である話者属性を、1以上格納し得る第三話者属性格納部と、2以上の言語のうちのすべての言語または2以上の一部の言語について、音声認識モデルを格納し得る音声認識モデル格納部と、音声情報を受信する音声情報受信部と、音声情報受信部が受信した音声情報を、音声認識モデル格納部の音声認識モデルを用いて音声認識し、音声認識結果を取得する音声認識部と、1以上の話者属性に応じて、2以上の翻訳サーバ装置のうちの翻訳サーバ装置を選択する翻訳サーバ選択部と、翻訳サーバ選択部が選択した翻訳サーバ装置に、音声認識結果を送信する音声認識結果送信部とを具備し、翻訳サーバ装置は、2以上

の言語のうちのすべての言語または2以上の一部の言語について、翻訳モデルを格納し得る翻訳モデル格納部と、音声認識結果を受信する音声認識結果受信部と、音声認識結果受信部が受信した音声認識結果を、翻訳モデル格納部の翻訳モデルを用いて、目的言語に翻訳し、翻訳結果を取得する翻訳部と、翻訳結果を送信する翻訳結果送信部とを具備し、音声合成サーバ装置は、2以上の言語のうちのすべての言語または2以上の一部の言語について、音声合成モデルを格納し得る音声合成モデル格納部と、翻訳結果を受信する翻訳結果受信部と、翻訳結果受信部が受信した翻訳結果を、音声合成モデル格納部の音声合成モデルを用いて音声合成し、音声合成結果を取得する音声合成部と、音声合成結果を第二端末装置に送信する音声合成結果送信部とを具備する音声翻訳システムである。

[0009] かかる構成により、音声翻訳において、話者の属性に応じて翻訳を行う装置を変更でき、翻訳処理において、精度が向上する。

[0010] また、本第四の発明の音声翻訳システムは、1以上の音声認識サーバ装置、1以上の翻訳サーバ装置、1以上の音声合成サーバ装置とを有する音声翻訳システムであって、音声認識サーバ装置は、2以上の言語のうちのすべての言語または2以上の一部の言語について、音声認識モデルを格納し得る音声認識モデル格納部と、音声情報を受信する音声情報受信部と、音声情報受信部が受信した音声情報を、音声認識モデル格納部の音声認識モデルを用いて音声認識し、音声認識結果を取得する音声認識部と、翻訳サーバ装置に、音声認識結果を送信する音声認識結果送信部とを具備し、翻訳サーバ装置は、2以上の言語のうちのすべての言語または2以上の一部の言語について、2以上の翻訳モデルを格納し得る翻訳モデル格納部と、1以上の話者属性を格納し得る第四話者属性格納部と、音声認識結果を受信する音声認識結果受信部と、1以上の話者属性に応じて、2以上の翻訳モデルから、一の翻訳モデルを選択する翻訳モデル選択部と、音声認識結果受信部が受信した音声認識結果を、翻訳モデル選択部が選択した翻訳モデルを用いて、目的言語に翻訳し、翻訳結果を取得する翻訳部と、翻訳結果を送信する翻訳結果送信部と

を具備し、音声合成サーバ装置は、2以上の言語のうちのすべての言語または2以上の一部の言語について、音声合成モデルを格納し得る音声合成モデル格納部と、翻訳結果を受信する翻訳結果受信部と、翻訳結果受信部が受信した翻訳結果を、音声合成モデル格納部の音声合成モデルを用いて音声合成し、音声合成結果を取得する音声合成部と、音声合成結果を第二端末装置に送信する音声合成結果送信部とを具備する音声翻訳システムである。

[0011] かかる構成により、音声翻訳において、話者の属性に応じて翻訳行うモデルを変更でき、翻訳処理において、精度が向上する。

[0012] また、本第五の発明の音声翻訳システムは、1以上の音声認識サーバ装置、1以上の翻訳サーバ装置、2以上の音声合成サーバ装置とを有する音声翻訳システムであって、音声認識サーバ装置は、2以上の言語のうちのすべての言語または2以上の一部の言語について、音声認識モデルを格納し得る音声認識モデル格納部と、音声情報を受信する音声情報受信部と、音声情報受信部が受信した音声情報を、音声認識モデル格納部の音声認識モデルを用いて音声認識し、音声認識結果を取得する音声認識部と、翻訳サーバ装置に、音声認識結果を送信する音声認識結果送信部とを具備し、翻訳サーバ装置は、2以上の言語のうちのすべての言語または2以上の一部の言語について、翻訳モデルを格納し得る翻訳モデル格納部と、1以上の話者属性を格納し得る第四話者属性格納部と、音声認識結果を受信する音声認識結果受信部と、音声認識結果受信部が受信した音声認識結果を、翻訳モデル格納部の翻訳モデルを用いて、目的言語に翻訳し、翻訳結果を取得する翻訳部と、1以上の話者属性に応じて、2以上の音声合成サーバ装置のうち一の音声合成サーバ装置を選択する音声合成サーバ選択部と、音声合成サーバ選択部が選択した音声合成サーバ装置に、翻訳結果を送信する翻訳結果送信部とを具備し、音声合成サーバ装置は、2以上の言語のうちのすべての言語または2以上の一部の言語について、音声合成モデルを格納し得る音声合成モデル格納部と、翻訳結果を受信する翻訳結果受信部と、翻訳結果受信部が受信した翻訳結果を、音声合成モデル格納部の音声合成モデルを用いて音声合成し、音声合成

結果を取得する音声合成部と、音声合成結果を第二端末装置に送信する音声合成結果送信部とを具備する音声翻訳システムである。

[0013] かかる構成により、音声翻訳において、話者の属性に応じて音声合成を行う装置を変更でき、音声合成処理において、適切な出力が行われる。

[0014] また、本第六の発明の音声翻訳システムは、1以上の音声認識サーバ装置、1以上の翻訳サーバ装置、1以上の音声合成サーバ装置とを有する音声翻訳システムであって、音声認識サーバ装置は、2以上の言語のうちのすべての言語または2以上の一部の言語について、音声認識モデルを格納し得る音声認識モデル格納部と、音声情報を受信する音声情報受信部と、音声情報受信部が受信した音声情報を、音声認識モデル格納部の音声認識モデルを用いて音声認識し、音声認識結果を取得する音声認識部と、翻訳サーバ装置に、音声認識結果を送信する音声認識結果送信部とを具備し、翻訳サーバ装置は、2以上の言語のうちのすべての言語または2以上の一部の言語について、翻訳モデルを格納し得る翻訳モデル格納部と、音声認識結果を受信する音声認識結果受信部と、音声認識結果受信部が受信した音声認識結果を、翻訳モデル格納部の翻訳モデルを用いて、目的言語に翻訳し、翻訳結果を取得する翻訳部と、音声合成サーバ装置に、翻訳結果を送信する翻訳結果送信部とを具備し、音声合成サーバ装置は、2以上の言語のうちのすべての言語または2以上の一部の言語について、2以上の音声合成モデルを格納し得る音声合成モデル格納部と、1以上の話者属性を格納し得る第五話者属性格納部と、翻訳結果を受信する翻訳結果受信部と、1以上の話者属性に応じて、2以上の音声合成モデルから、一の音声合成モデルを選択する音声合成モデル選択部と、翻訳結果受信部が受信した翻訳結果を、音声合成モデル選択部が選択した音声合成モデルを用いて音声合成し、音声合成結果を取得する音声合成部と、音声合成結果を第二端末装置に送信する音声合成結果送信部とを具備する音声翻訳システムである。

[0015] かかる構成により、音声翻訳において、話者の属性に応じて音声合成を行うモデルを変更でき、音声合成処理において、適切な出力が行われる。

[0016] また、本第七の発明の音声翻訳システムは、第一の発明に対して、第一端末装置は、１以上の話者属性を受け付ける第一話者属性受付部と、１以上の話者属性を、第一話者属性格納部に蓄積する第一話者属性蓄積部とを具備する音声翻訳システムである。

[0017] かかる構成により、話者が用いる端末において、話者属性の設定ができるために、音声認識や翻訳や音声合成の各処理において話者属性が利用でき、音声認識や翻訳や音声合成の各処理において、精度が向上し、適切な出力が行われる。

[0018] また、本第八の発明の音声翻訳システムは、第二から第六いずれかの発明に対して、音声認識サーバ装置は、音声情報受信部が受信した音声情報から、１以上の音声に関する話者属性を取得する音声話者属性取得部と、音声話者属性取得部が取得した１以上の話者属性を、第三話者属性格納部に蓄積する第三話者属性蓄積部とをさらに具備する音声翻訳システムである。

[0019] かかる構成により、自動的に音声に関する話者属性が取得できるために、音声認識や翻訳や音声合成の各処理において話者属性が利用でき、音声認識や翻訳や音声合成の各処理において、精度が向上し、適切な出力が行われる。

[0020] また、本第九の発明の音声翻訳システムは、第二から第六いずれかの発明に対して、翻訳サーバ装置は、音声認識結果受信部が受信した音声認識結果から、１以上の言語に関する話者属性を取得する言語話者属性取得部と、言語話者属性取得部が取得した１以上の話者属性を、第四話者属性格納部に蓄積する第四話者属性蓄積部とをさらに具備する音声翻訳システムである。

[0021] かかる構成により、自動的に言語に関する話者属性が取得できるために、音声認識や翻訳や音声合成の各処理において話者属性が利用でき、音声認識や翻訳や音声合成の各処理において、精度が向上し、適切な出力が行われる。

[0022] また、本第十の発明の音声翻訳システムは、第一から第九いずれかの発明に対して、話者が使用する言語である原言語を特定する原言語識別子、およ

び翻訳先の言語である目的言語を特定する目的言語識別子、および 1 以上の話者属性を含む音声翻訳制御情報が、音声認識サーバ装置から 1 以上の翻訳サーバ装置を経由して、音声合成サーバ装置に送信され、音声認識サーバ選択部、または音声認識部、または音声認識モデル選択部、または翻訳サーバ選択部、または翻訳部、または翻訳モデル選択部、音声合成サーバ選択部、または音声合成部、または音声合成モデル選択部は、音声翻訳制御情報を用いて、各々の処理を行う音声翻訳システムである。

- [0023] かかる構成により、話者属性を含む音声翻訳制御情報が流通し、音声認識、翻訳、および音声合成において、同一の話者を想定した処理が行われ、音声認識や翻訳や音声合成の各処理において、精度が向上し、適切な出力が行われる。

発明の効果

- [0024] 本発明による音声翻訳システムによれば、音声翻訳において、音声認識、または翻訳、または音声合成の各処理において、精度が向上し、適切な出力が行われる。

発明を実施するための最良の形態

- [0025] 以下、音声翻訳システム等の実施形態について図面を参照して説明する。なお、実施の形態において同じ符号を付した構成要素は同様の動作を行うので、再度の説明を省略する場合がある。

(実施の形態 1)

- [0026] 本実施の形態において、ネットワーク型音声翻訳システムにおいて、話者属性に基づいて、音声認識を行うサーバ装置や音声認識モデルを適切に変更したり、翻訳を行うサーバ装置や翻訳モデルを適切に変更したり、音声合成を行うサーバ装置や音声合成モデルを適切に変更したりすることにより、精度の高い音声翻訳を行えるネットワーク型音声翻訳システムについて説明する。

- [0027] 図 1 は、本実施の形態における音声翻訳システムの概念図である。音声翻訳システムは、1 以上の第一端末装置 1、1 以上の第二端末装置 2、1 以上

の音声認識サーバ装置 3、1 以上の翻訳サーバ装置 4、1 以上の音声合成サーバ装置 5 を具備する。

[0028] 音声翻訳システムにおいて、例えば、第一端末装置 1 のユーザ（ユーザ A）である日本人が日本語で「おはようございます」と発話した場合、音声認識サーバ装置 3 は日本語「おはようございます」を音声認識する。そして、翻訳サーバ装置 4 は音声認識結果を、例えば、英語「Good morning」に翻訳する。次に、音声合成サーバ装置 5 は英文「Good morning」から「Good morning」の音声情報を作り出す。そして、英語のネイティブであるユーザ B の第二端末装置 2 から、音声「Good morning」が出力される。

[0029] 第一端末装置 1 および第二端末装置 2 は、例えば、通話する端末（電話、携帯電話を含む）である。ここでは、主として、第一端末装置 1 を発話する側の端末、第二端末装置 2 を発話される側の端末として説明するが、両者は入れ替われることは言うまでもない。また、通常、第一端末装置 1 と第二端末装置 2 とが、発話する側の端末、発話される側の端末として、順次、連続的に入れ替わりながら、第一端末装置 1 のユーザ A と第二端末装置 2 のユーザ B とが会話を進めていく。また、第一端末装置 1 および第二端末装置 2 は、同様の機能（下記の構成要素）を有するとして説明するが、それぞれ、一部に有しない機能があっても良いことは言うまでもない。また、第一端末装置 1 のユーザ（話者）をユーザ A、第二端末装置 2 のユーザ（話者）をユーザ B として、以下、説明する。

[0030] 図 2 は、本実施の形態における音声翻訳システムのブロック図である。また、図 3 は、音声認識サーバ装置 3 のブロック図である。図 4 は、翻訳サーバ装置 4 のブロック図である。図 5 は、音声合成サーバ装置 5 のブロック図である。

[0031] 音声翻訳システムは、1 以上の第一端末装置 1、1 以上の第二端末装置 2、1 以上の音声認識サーバ装置 3、1 以上の翻訳サーバ装置 4、1 以上の音声合成サーバ装置 5 を備える。

- [0032] 第一端末装置 1 は、第一話者属性格納部 1 1、第一話者属性受付部 1 2、第一話者属性蓄積部 1 3、第一音声受付部 1 4、第一音声認識サーバ選択部 1 5、第一音声送信部 1 6、第一音声受信部 1 7、第一音声出力部 1 8、第一話者属性送信部 1 9 を備える。第一音声認識サーバ選択部 1 5 は、第一サーバ選択情報格納手段 1 5 1、第一サーバ選択手段 1 5 2 を備える。
- [0033] 第二端末装置 2 は、第二話者属性格納部 2 1、第二話者属性受付部 2 2、第二話者属性蓄積部 2 3、第二音声受付部 2 4、第二音声認識サーバ選択部 2 5、第二音声送信部 2 6、第二音声受信部 2 7、第二音声出力部 2 8、第二話者属性送信部 2 9 を備える。第二音声認識サーバ選択部 2 5 は、第二サーバ選択情報格納手段 2 5 1、第二サーバ選択手段 2 5 2 を備える。
- [0034] 音声認識サーバ装置 3 は、第三話者属性格納部 3 0 1、音声認識モデル格納部 3 0 2、第三話者属性受信部 3 0 3、音声話者属性取得部 3 0 4、第三話者属性蓄積部 3 0 5、音声情報受信部 3 0 6、音声認識モデル選択部 3 0 7、音声認識部 3 0 8、翻訳サーバ選択部 3 0 9、音声認識結果送信部 3 1 0、第三話者属性送信部 3 1 1 を備える。音声認識モデル選択部 3 0 7 は、第三モデル選択情報格納手段 3 0 7 1、第三モデル選択手段 3 0 7 2 を備える。また、翻訳サーバ選択部 3 0 9 は、第三サーバ選択情報格納手段 3 0 9 1、第三サーバ選択手段 3 0 9 2 を備える。
- [0035] 翻訳サーバ装置 4 は、第四話者属性格納部 4 0 1、翻訳モデル格納部 4 0 2、第四話者属性受信部 4 0 3、言語話者属性取得部 4 0 4、第四話者属性蓄積部 4 0 5、音声認識結果受信部 4 0 6、翻訳モデル選択部 4 0 7、翻訳部 4 0 8、音声合成サーバ選択部 4 0 9、翻訳結果送信部 4 1 0、第四話者属性送信部 4 1 1 を備える。翻訳モデル選択部 4 0 7 は、第四モデル選択情報格納手段 4 0 7 1、第四モデル選択手段 4 0 7 2 を備える。音声合成サーバ選択部 4 0 9 は、第四サーバ選択情報格納手段 4 0 9 1、第四サーバ選択手段 4 0 9 2 を備える。
- [0036] 音声合成サーバ装置 5 は、第五話者属性格納部 5 0 1、音声合成モデル格納部 5 0 2、第五話者属性受信部 5 0 3、第五話者属性蓄積部 5 0 4、翻訳

結果受信部 505、音声合成モデル選択部 506、音声合成部 507、音声合成結果送信部 508 を備える。音声合成モデル選択部 506 は、第五モデル選択情報格納手段 5061、第五モデル選択手段 5062 を備える。

[0037] 以下、各構成要素の機能、実現手段等について説明する。ただし、第一端末装置 1 の構成要素の機能と、対応する第二端末装置 2 の構成要素の機能とは同じ（例えば、第一話者属性受付部 12 と第二話者属性受付部 22 の機能は同じ）であるので、第一端末装置 1 の構成要素についてのみ説明する。

[0038] 第一端末装置 1 を構成する第一話者属性格納部 11 は、1 以上の話者属性を格納し得る。話者属性とは、話者の属性値である。話者属性とは、例えば、話者の性別、年齢、話速、話者クラス（話者が使用する言語に関して、使う単語の難しさ、文法の正しさなどを考慮した熟練度）などである。話者クラスは、使う単語の難しさの度合いを示す情報、使う用語の丁寧さの度合いを示す情報、文法の正しさの度合いを示す情報、およびこれらの複合的な度合いを示す情報、ネイティブか否かを示す情報などである。また、話者属性には、話者の感情（嬉しい、悲しいなど）なども含まれても良い。なお、話者属性の内容は問わないことは言うまでもない。また、話者属性には、話者が特定されれば、一時的には変化の無い属性である静的話者属性情報と、時々により変化し得る属性である動的話者属性情報とがある。静的話者属性情報には、例えば、上述した話者の性別、年齢、話者クラスなどがある。また、動的話者属性情報には、話速や感情などがある。

第一話者属性格納部 11 に格納されている話者属性は、通常、第一端末装置 1 のユーザが入力した情報である。また、第一話者属性格納部 11 に格納されている話者属性は、通常、静的話者属性情報である。第一話者属性格納部 11 は、1 以上の話者属性を含む音声翻訳制御情報を格納していても良い。かかる場合、第一話者属性格納部 11 は、第一音声翻訳制御情報格納部 11 と呼んでも良い。

[0039] 第一話者属性受付部 12 は、第一端末装置 1 のユーザから、1 以上の話者属性を受け付ける。ここで、受け付けとは、キーボードやマウス、タッチパ

ネルなどの入力デバイスから入力された情報の受け付け、有線もしくは無線の通信回線を介して送信された情報の受信、光ディスクや磁気ディスク、半導体メモリなどの記録媒体から読み出された情報の受け付けなどを含む概念である。話者属性の入力手段は、テンキーやキーボードやマウスやメニュー画面によるもの等、何でも良い。

[0040] 第一話者属性蓄積部 13 は、第一話者属性受付部 12 が受け付けた 1 以上の話者属性を、第一話者属性格納部 11 に蓄積する。

[0041] 第一音声受付部 14 は、第一端末装置 1 のユーザ（ユーザ A という）から音声を受け付ける。

[0042] 第一音声認識サーバ選択部 15 は、1 以上の話者属性に応じて、2 以上の音声認識サーバ装置 3 のうちの音声認識サーバ装置 3 を選択する。例えば、第一音声認識サーバ選択部 15 は、音声認識サーバ選択情報管理表を用いて選択する。第一音声認識サーバ選択部 15 は、図示しないサーバ装置（後述する制御装置 256）にあっても良い。かかる場合、第一音声送信部 16 も図示しないサーバ装置に存在しても良い。ここで、音声認識サーバ装置 3 の選択とは、例えば、一の音声認識サーバ装置 3 と通信するための情報（例えば、一の音声認識サーバ装置 3 の IP アドレス）などを取得する処理である。また、第一音声認識サーバ選択部 15 は、1 以上の話者属性を含む音声翻訳制御情報を用いて、2 以上の音声認識サーバ装置 3 のうちの音声認識サーバ装置 3 を選択することは好適である。音声翻訳制御情報とは、音声認識サーバ装置 3、翻訳サーバ装置 4、および音声合成サーバ装置 5 が、それぞれ音声認識、翻訳、および音声合成を行ったり、処理結果を送信したりするための情報を有する。音声翻訳制御情報は、例えば、処理結果を送信する先を特定する情報（IP アドレスや電話番号など）、原言語や目的言語を特定する情報（日本語、英語、ドイツ語など）などを有する。第一端末装置 1（第二端末装置 2 も同様である）の図示しない受付部（第一話者属性受付部 12 でも良い）は、例えば、原言語と目的言語を、ユーザから受け付ける。また、第一音声認識サーバ選択部 15 は、例えば、第二端末装置 2 の電話番

号やIPアドレスなどから、自動的に原言語と目的言語を決定する。かかる場合、第一音声認識サーバ選択部15は、電話番号やIPアドレスなどの情報と、言語を特定する情報を対応付けて保持していたり、電話番号やIPアドレスなどの情報をキーとして他の装置から言語を特定する情報を取得したりする。また、第一音声認識サーバ選択部15は、原言語や目的言語から音声認識すべき音声認識サーバ装置3を特定する情報（IPアドレスなど）や、翻訳サーバ装置4を特定する情報（IPアドレスなど）や、音声合成サーバ装置5を特定する情報（IPアドレスなど）を取得する。つまり、第一音声認識サーバ選択部15は、原言語や目的言語と、各サーバ装置を特定する情報とを対応付けて有していたり、原言語や目的言語を用いて、他の装置から、各サーバ装置を特定する情報を取得したりする。また、音声翻訳制御情報は、通常、1以上の話者属性を有し、かつ、入力音声のフォーマットを示す情報、出力音声のフォーマットを示す情報、入出力音声の声質を指定する情報、入力テキストのフォーマットを示す情報、出力テキストのフォーマットを示す情報なども有しても良い。

[0043] 第一サーバ選択情報格納手段151は、音声認識サーバ装置3を識別する音声認識サーバ識別子と、1以上の話者属性とを対応付けて有する音声認識サーバ選択情報を、2以上有する音声認識サーバ選択情報管理表を格納している。音声認識サーバ識別子は、例えば、音声認識サーバ装置3と通信するための情報（例えば、一の音声認識サーバ装置3のIPアドレス）である。

[0044] 第一サーバ選択手段152は、第一話者属性格納部11に格納されている1以上の話者属性に対応する音声認識サーバ識別子を、第一サーバ選択情報格納手段151から検索する。

[0045] 第一音声送信部16は、第一音声受付部14が受け付けた音声から構成した音声情報を送信する。音声の送信先は、1以上の音声認識サーバ装置3のうちのいずれかである。第一音声送信部16は、第一音声認識サーバ選択部15が選択した音声認識サーバ装置3に、第一音声受付部14が受け付けた音声から構成される音声情報を送信することは好適である。なお、第一音声

送信部 16 は、2 以上の音声認識サーバ装置 3 に音声情報を送信しても良い。また、ここで、音声情報は、音声から構成されるデジタル情報であることは好適である。さらに、第一音声送信部 16 は、直接的に音声を 1 以上の音声認識サーバ装置 3 に送信しても良いし、他の装置経由で（間接的に）、音声を 1 以上の音声認識サーバ装置 3 に送信しても良い。

[0046] 第一音声受信部 17 は、第二端末装置 2 のユーザ B から発声された音声から構成された音声情報を音声翻訳した結果である音声情報を受信する。この音声情報は、通常、第一端末装置 1 のユーザ A が理解できる言語に翻訳された音声の情報である。この音声情報は、通常、第二端末装置 2 から音声認識サーバ装置 3、翻訳サーバ装置 4、および音声合成サーバ装置 5 を経由して送信されてきた情報である。

[0047] 第一音声出力部 18 は、第一音声受信部 17 が受信した音声情報を出力する。ここで、出力とは、通常、スピーカへの音声出力である。

[0048] 第一話者属性送信部 19 は、第一話者属性格納部 11 に格納されている 1 以上の話者属性を、第一音声認識サーバ選択部 15 が選択した音声認識サーバ装置 3 に送信する。第一話者属性送信部 19 は、1 以上の話者属性を含む音声翻訳制御情報を、第一音声認識サーバ選択部 15 が選択した音声認識サーバ装置 3 に送信しても良い。かかる場合、第一話者属性送信部 19 は、第一音声翻訳制御情報送信部 19 と言い換えても良い。

[0049] 音声認識サーバ装置 3 を構成する第三話者属性格納部 301 は、話者の属性値である話者属性を、1 以上格納し得る。ここでの話者属性は、音声認識サーバ装置 3 で取得した話者属性でも良いし、第一端末装置 1 や第二端末装置 2 から送信された話者属性でも良いし、音声認識サーバ装置 3 で取得した話者属性および第一端末装置 1 や第二端末装置 2 から送信された話者属性の両方を含むものでも良い。また、第三話者属性格納部 301 は、1 以上の話者属性を含む音声翻訳制御情報を格納していても良い。第三話者属性格納部 301 は、第三音声翻訳制御情報格納部 301 と呼んでも良い。

[0050] 音声認識モデル格納部 302 は、2 以上の言語のうちのすべての言語また

は2以上の一部の言語について、音声認識モデルを格納し得る。音声認識モデル格納部302は、2以上の音声認識モデルを格納しても良い。音声認識モデルとは、例えば、隠れマルコフモデル（HMM）の音響モデルである。ただし、音声認識モデルは、必ずしも、HMMの音響モデルである必要はない。音声認識モデルは、単一ガウス分布モデルや、確率モデル（GMM：ガウシアンミクスチャモデル）や、統計モデルなど、他のモデルに基づく音響モデルでも良い。

[0051] 第三話者属性受信部303は、第一端末装置1から、直接的または間接的に、1以上の話者属性を受信する。第三話者属性受信部303は、第一端末装置1から、音声翻訳制御情報を受信しても良い。かかる場合、第三話者属性受信部303は、第三音声翻訳制御情報受信部303と呼んでも良い。

[0052] 音声話者属性取得部304は、音声情報受信部306が受信した音声情報から、1以上の音声話者属性を取得する。音声話者属性とは、音声に関する話者属性であり、音声から取得できる話者属性である。音声話者属性取得部304は、動的話者属性情報を取得することは好適である。また、音声話者属性取得部304は、静的話者属性情報を取得しても良い。

また、音声話者属性取得部304は、例えば、音声情報をスペクトル分析し、1以上の特徴量を取得する。そして、音声話者属性取得部304は、1以上の特徴量から、話者の年齢、性別、話速、感情などの話者属性を決定する。音声話者属性取得部304は、例えば、男性または／および女性であることを決定するための特徴量の情報（特徴量をパラメータとする条件）を保持しており、取得した1以上の特徴量から、話者が男性か女性かを決定して、性別の情報（例えば、男性「0」、女性「1」）を取得する。また、音声話者属性取得部304は、例えば、特定の年齢、または特定の年齢層（例えば、10代、20代など）を決定するための特徴量の情報を保持しており、取得した1以上の特徴量から、話者の年齢または年齢層を決定して、年齢または年齢層の情報（例えば、9歳まで「0」、10代「1」など）を取得する。また、音声話者属性取得部304は、音声情報を解析し、話速（例えば

、4. 5音／秒）を取得する。話速を取得する技術は公知技術であるので、詳細な説明を省略する。また、音声話者属性取得部304は、例えば、取得した1以上の特徴量から感情（動的話者属性情報の一種）を取得しても良い。さらに具体的には、音声話者属性取得部304は、例えば、感情「普通」の場合のピッチとパワーの値を保持している。そして、音声話者属性取得部304は、抽出した有声部分のピッチとパワーの値から平均値・最大値・最小値を求める。そして、音声話者属性取得部304は、感情「普通」の場合のピッチとパワーの値と、抽出した有声部分のピッチとパワーの平均値・最大値・最小値とを用いて、平均ピッチが低く、平均パワーが高い場合は、感情「怒り」を取得する。また、音声話者属性取得部304は、感情「普通」の場合のピッチとパワーの値と比較して、最小ピッチが高く、最大パワーが低い場合、感情「悲しみ」を取得する。また、感情「普通」の場合のピッチとパワーの値と比較して、特徴量が大きい場合、音声話者属性取得部304は、感情「喜び」を取得する。

音声話者属性取得部304は、取得した1以上の特徴量のうち、パワーおよび韻律を用いて感情を取得することは好適である。感情を取得する手法については、URL「<http://www.kansei.soft.iwate-pu.ac.jp/abstract/2007/0312004126.pdf>」の論文を参照のこと。

なお、音声話者属性取得部304が、話速等の属性を取得する音声情報の単位は問わない。つまり、音声話者属性取得部304は、文の単位で話速等の属性を取得しても良いし、単語の単位で話速等の属性を取得しても良いし、認識結果の単位で話速等の属性を取得しても良いし、複数の文の単位で話速等の属性を取得しても良い。

[0053] また、音声話者属性取得部304は、音声認識部308が音声認識した結果である音声認識結果から、話者クラス（この話者クラスは、言語話者属性の一種である）を決定しても良い。例えば、音声話者属性取得部304は、難解度を有する用語辞書（用語と難解度とを対応づけて有する2以上の用語情報の集合）を保持しており、音声認識結果に含まれる1以上の用語の難解

度（ $n_1, n_2, \dots$ ）を取得し、当該1以上の難解度から話者クラス（難解度が高「0」、難解度が中「1」、難解度が低「2」など）を決定する。また、音声話者属性取得部304は、1以上の用語の難解度（ $n_1, n_2, \dots$ ）、および文法ミスの有無を用いて話者クラスを決定する。例えば、音声話者属性取得部304は、文法ミスが存在する場合には、1以上の難解度から取得される最終難解度（高「0」、難解度が中「1」、難解度が低「2」）に「1」を加えた値を話者クラスとして取得する。なお、文の中に文法ミスが存在するか否かの検査処理は、公知の自然言語処理であるので、詳細な説明は省略する。なお、音声話者属性取得部304は、上述した方法以外の方法により、話者属性を取得しても良いし、いかなる話者属性を取得しても良い。また、音声話者属性取得部304に関する技術は、例えば、「不特定話者の音声自動認識のための性別・年齢差による話者分類の考察」（中川聖一他，電子通信学会論文誌）に開示されている（<http://www.slp.ics.tut.ac.jp/shiryou/number-1/J1980-06.pdf>参照）。また、音声話者属性取得部304は、機械学習を用いて、性別や年齢などの話者属性を決定しても良い。つまり、音声話者属性取得部304は、性別（男性または女性）と1以上の特徴量とを有する情報の組を複数組格納しており、音声情報受信部306が受信した音声情報から得られた1以上の特徴量に対して、SVMや決定木などの機械学習のアルゴリズムにより、音声情報受信部306が受信した音声情報に対応する話者の性別を決定しても良い。

[0054] 第三話者属性蓄積部305は、音声話者属性取得部304が取得した1以上の話者属性を、第三話者属性格納部301に蓄積する。なお、ここでの蓄積は、一時的な蓄積でも良い。

[0055] 音声情報受信部306は、第一端末装置1から、直接的にまたは間接的に音声情報を受信する。

[0056] 音声認識モデル選択部307は、1以上の話者属性に応じて、2以上の音声認識モデルから、一の音声認識モデルを選択する。

[0057] 第三モデル選択情報格納手段3071は、音声認識モデル選択情報管理表

を格納している。音声認識モデル選択情報管理表は、１以上の話者属性に関する条件（１以上の話者属性でも良い）と、音声認識モデルを識別する音声認識モデル識別子とを有するレコードを１以上含む表である。

[0058] 第三モデル選択手段３０７２は、第三話者属性格納部３０１に格納されている１以上の話者属性から音声認識モデル選択情報管理表を検索し、１以上の話者属性に対応する音声認識モデル識別子を取得する。なお、音声認識部３０８は、第三モデル選択手段３０７２が取得した音声認識モデル識別子で識別される音声認識モデルを、音声認識モデル格納部３０２から読み出し、当該音声認識モデルを用いて、音声認識処理を行う。

[0059] 音声認識部３０８は、音声情報受信部３０６が受信した音声情報を、音声認識モデル格納部３０２の音声認識モデルを用いて音声認識し、音声認識結果を取得する。また、音声認識部３０８は、音声情報受信部３０６が受信した音声情報を、音声認識モデル選択部３０７が選択した音声認識モデルを用いて音声認識し、音声認識結果を取得することは好適である。音声認識部３０８は、いかなる音声認識方法でも良い。音声認識部３０８は公知技術である。また、音声認識の対象の言語（原言語）の情報は、例えば、音声翻訳制御情報の中に含まれる。例えば、音声翻訳制御情報は、第一端末装置、音声認識サーバ装置、翻訳サーバ装置、音声合成サーバ装置、および第二端末装置２の間を転送されていく。音声認識結果は、通常、原言語（第一端末装置１のユーザＡが発話した音声の言語）の文字列である。

[0060] 翻訳サーバ選択部３０９は、１以上の話者属性に応じて、２以上の翻訳サーバ装置４のうちの翻訳サーバ装置４を選択する。なお、翻訳サーバ選択部３０９は、図示しないサーバ装置にあっても良い。かかる場合、通常、音声認識結果送信部３１０も図示しないサーバ装置に存在する。また、かかる場合、選択すべき翻訳サーバ装置４の情報が図示しないサーバ装置から送信されても良い。

[0061] 第三サーバ選択情報格納手段３０９１は、翻訳サーバ選択情報管理表を格納している。翻訳サーバ選択情報管理表は、１以上の話者属性に関する条件

(1以上の話者属性でも良い)と、翻訳サーバ装置4を識別する翻訳サーバ装置識別子とを有するレコードを1以上含む表である。

[0062] 第三サーバ選択手段3092は、第三話者属性格納部301に格納されている1以上の話者属性から翻訳サーバ選択情報管理表を検索し、1以上の話者属性に対応する翻訳サーバ装置識別子を取得する。なお、音声認識結果送信部310は、第三サーバ選択手段3092が取得した翻訳サーバ装置識別子に対応する翻訳サーバ装置4に、音声認識結果を送信する。

[0063] 音声認識結果送信部310は、翻訳サーバ装置4に、直接的にまたは間接的に音声認識結果を送信する。音声認識結果送信部310は、翻訳サーバ選択部309が選択した一の翻訳サーバ装置4に、直接的にまたは間接的に音声認識結果を送信することは好適である。

[0064] 第三話者属性送信部311は、第三話者属性格納部301に格納されている1以上の話者属性を、直接的にまたは間接的に翻訳サーバ装置4に送信する。第三話者属性送信部311は、音声翻訳制御情報を翻訳サーバ装置4に送信しても良い。かかる場合、第三話者属性送信部311は、第三音声翻訳制御情報送信部311と呼んでも良い。

[0065] 翻訳サーバ装置4を構成する第四話者属性格納部401は、1以上の話者属性を格納し得る。第四話者属性格納部401は、音声翻訳制御情報を格納していても良い。かかる場合、第四話者属性格納部401は、第四音声翻訳制御情報格納部401と呼んでも良い。

[0066] 翻訳モデル格納部402は、2以上の言語のうちのすべての言語または2以上の一部の言語について、翻訳用情報を格納し得る。翻訳モデル格納部402は、2以上の翻訳用情報を格納していても良い。翻訳用情報は、例えば、翻訳モデルと言語モデルである。また、翻訳モデル格納部402は、言語モデルを有さず、1以上の翻訳モデルのみを格納していても良い。かかる場合、翻訳モデル格納部402は、1以上の翻訳モデルを格納し得る。なお、後述する翻訳モデルの選択は、翻訳用情報の選択と同意義である、とする。

[0067] 第四話者属性受信部403は、1以上の話者属性を、音声認識サーバ装置

3から直接にまたは間接的に受信する。第四話者属性受信部403は、音声翻訳制御情報を音声認識サーバ装置3から受信しても良い。かかる場合、第四話者属性受信部403は、第四音声翻訳制御情報受信部403と呼んでも良い。

[0068] 言語話者属性取得部404は、音声認識結果受信部406が受信した音声認識結果から、1以上の言語話者属性を取得する。言語話者属性とは、言語処理によって取得され得る話者属性である。言語話者属性取得部404は、例えば、音声認識結果を自然言語処理に、話者クラスを特定する。話者クラスとは、例えば、使う単語の難しさ、文法の正しさなどを考慮した言語の熟練度により、話者をカテゴライズした情報である。言語話者属性取得部404は、例えば、音声認識結果の中で、難解な用語が使われている頻度や割合から、話者クラスを決定する。また、言語話者属性取得部404は、例えば、音声認識結果の中で、丁寧語が使われているか否か、丁寧語が使われている頻度や割合から、話者クラスを決定する。また、言語話者属性取得部404は、例えば、音声認識結果の中で、女子高生が多用する新語が使われているか否か、新語が使われている頻度や割合から、話者クラスを決定する。言語話者属性取得部404は、例えば、難解な用語や丁寧語を構成する形態素や女子高生が多用する新語などを格納しており、音声認識結果を形態素解析し、難解な用語や丁寧語を構成する形態素や女子高生が多用する新語などが存在する頻度や割合を取得し、予め決められた条件（ランクA：難解な用語が10%以上、B：難解な用語が1%以上10%未満、C：難解な用語が1%未満など）により、話者クラス（ランク）を決定する。言語話者属性取得部404は、その他のアルゴリズムにより、音声認識結果を自然言語処理し、1以上の言語話者属性を取得しても良い。なお、言語話者属性取得部404は、音声話者属性取得部304における言語話者属性の取得方法と同様の方法により、1以上の言語話者属性を取得しても良い。

[0069] 第四話者属性蓄積部405は、第四話者属性受信部403が受信した以上の話者属性を、第四話者属性格納部401に、少なくとも一時的に蓄積する

。第四話者属性蓄積部 405 は、音声翻訳制御情報を第四話者属性格納部 401 に蓄積しても良い。かかる場合、第四話者属性蓄積部 405 は、第四音声翻訳制御情報蓄積部 405 と呼んでも良い。

[0070] 音声認識結果受信部 406 は、音声認識結果を、音声認識サーバ装置 3 から直接、または間接的に受信する。

[0071] 翻訳モデル選択部 407 は、第四話者属性受信部 403 が受信した 1 以上の話者属性に応じて、2 以上の翻訳モデルから、一の翻訳モデルを選択する。

[0072] 第四モデル選択情報格納手段 4071 は、翻訳モデル選択情報管理表を格納している。翻訳モデル選択情報管理表は、1 以上の話者属性に関する条件（1 以上の話者属性でも良い）と、翻訳モデルを識別する翻訳モデル識別子とを有するレコードを 1 以上含む表である。

[0073] 第四モデル選択手段 4072 は、第四話者属性格納部 401 に格納されている 1 以上の話者属性から翻訳モデル選択情報管理表を検索し、1 以上の話者属性に対応する翻訳モデル装置識別子を取得する。なお、翻訳部 408 は、第四モデル選択手段 4072 が取得した翻訳モデル識別子に対応する翻訳モデルを翻訳モデル格納部 402 から取得し、当該翻訳モデルを用いて翻訳処理を行う。

[0074] 翻訳部 408 は、音声認識結果受信部 406 が受信した音声認識結果を、翻訳モデル格納部 402 の翻訳モデルを用いて、目的言語に翻訳し、翻訳結果を取得する。翻訳部 408 は、音声認識結果受信部 406 が受信した音声認識結果を、翻訳モデル選択部 407 が選択した翻訳モデルを用いて、目的言語に翻訳し、翻訳結果を取得することは好適である。なお、原言語と目的言語を特定する情報は、例えば、音声翻訳制御情報の中に含まれる。また、翻訳部 408 における翻訳方法と問わない。翻訳部 408 は公知技術である。

[0075] 音声合成サーバ選択部 409 は、1 以上の話者属性に応じて、2 以上の音声合成サーバ装置 5 のうちの音声合成サーバ装置 5 を選択する。

- [0076] 第四サーバ選択情報格納手段4091は、音声合成サーバ選択情報管理表を格納している。音声合成サーバ選択情報管理表は、1以上の話者属性に関する条件（1以上の話者属性でも良い）と、音声合成サーバ装置5を識別する音声合成装置識別子とを有するレコードを1以上含む表である。
- [0077] 第四サーバ選択手段4092は、第四話者属性格納部401に格納されている1以上の話者属性から音声合成サーバ選択情報管理表を検索し、1以上の話者属性に対応する音声合成サーバ装置識別子を取得する。なお、翻訳結果送信部410は、第四サーバ選択手段4092が取得した音声合成サーバ装置識別子に対応する音声合成サーバ装置5に、翻訳結果を送信する。
- [0078] 翻訳結果送信部410は、翻訳部408が翻訳処理を行った結果である翻訳結果を、音声合成サーバ装置5に、直接または間接的に送信する。また、翻訳結果送信部410は、音声合成サーバ選択部409はが選択した音声合成サーバ装置5に、翻訳結果を、直接または間接的に送信することは好適である。
- [0079] 第四話者属性送信部411は、音声合成サーバ装置5に、直接または間接的に、第四話者属性受信部403が受信した1以上の話者属性を送信する。第四話者属性送信部411は、音声翻訳制御情報を音声合成サーバ装置5に送信しても良い。かかる場合、第四話者属性送信部411は、第四音声翻訳制御情報送信部411と呼んでも良い。
- [0080] 音声合成サーバ装置5を構成する第五話者属性格納部501は、1以上の話者属性を格納し得る。第五話者属性格納部501は、音声翻訳制御情報を格納していても良い。かかる場合、第五話者属性格納部501は、第五音声翻訳制御情報格納部501と呼んでも良い。
- [0081] 音声合成モデル格納部502は、2以上の言語のうちのすべての言語または2以上の一部の言語について、音声合成モデルを格納し得る。音声合成モデル格納部502は、2以上の音声合成モデルを格納していても良い。
- [0082] 第五話者属性受信部503は、1以上の話者属性を、音声認識サーバ装置3から直接にまたは間接的に受信する。第五話者属性受信部503は、音声

翻訳制御情報を音声認識サーバ装置 3 から受信しても良い。かかる場合、第五話者属性受信部 503 は、第五音声翻訳制御情報受信部 503 と呼んでも良い。

[0083] 第五話者属性蓄積部 504 は、第五話者属性受信部 503 が受信した 1 以上の話者属性を、第五話者属性格納部 501 に、少なくとも一時的に蓄積する。第五話者属性蓄積部 504 は、音声翻訳制御情報を第五話者属性格納部 501 に蓄積しても良い。かかる場合、第五話者属性蓄積部 504 は、第五音声翻訳制御情報蓄積部 504 と呼んでも良い。

[0084] 翻訳結果受信部 505 は、翻訳サーバ装置 4 から直接または間接的に、翻訳結果を受信する。

[0085] 音声合成モデル選択部 506 は、第五話者属性受信部 503 が受信した 1 以上の話者属性に応じて、2 以上の音声合成モデルから、一の音声合成モデルを選択する。

[0086] 第五モデル選択情報格納手段 5061 は、音声合成モデル選択情報管理表を格納している。音声合成モデル選択情報管理表は、1 以上の話者属性に関する条件（1 以上の話者属性でも良い）と、音声合成モデルを識別する音声合成モデル識別子とを有するレコードを 1 以上含む表である。

[0087] 第五モデル選択手段 5062 は、第五話者属性格納部 501 に格納されている 1 以上の話者属性から音声合成モデル選択情報管理表を検索し、1 以上の話者属性に対応する音声合成モデル装置識別子を取得する。なお、音声合成部 507 は、第五モデル選択手段 5062 が取得した音声合成モデル識別子に対応する音声合成モデルを音声合成モデル格納部 502 から取得し、当該音声合成モデルを用いて音声合成処理を行う。

[0088] 音声合成部 507 は、翻訳結果受信部 505 が受信した翻訳結果を、音声合成モデル格納部 502 の音声合成モデルを用いて音声合成し、音声合成結果を取得する。音声合成部 507 は、翻訳結果受信部 505 が受信した翻訳結果を、音声合成モデル選択部 506 が選択した音声合成モデルを用いて音声合成し、音声合成結果を取得することは好適である。ここで、音声合成を

する目的言語を特定する情報は、例えば、音声翻訳制御情報の中に含まれる。
。

[0089] 音声合成結果送信部 508 は、音声合成部 507 が取得した音声合成結果を、直接または間接的に第二端末装置 2 に送信する。

[0090] 第一話者属性格納部 11、第一サーバ選択情報格納手段 151、第二話者属性格納部 21、第二サーバ選択情報格納手段 251、第三話者属性格納部 301、音声認識モデル格納部 302、第三モデル選択情報格納手段 3071、第三サーバ選択情報格納手段 3091、第四話者属性格納部 401、翻訳モデル格納部 402、第四モデル選択情報格納手段 4071、第四サーバ選択情報格納手段 4091、第五話者属性格納部 501、音声合成モデル格納部 502、および第五モデル選択情報格納手段 5061 は、不揮発性の記録媒体が好適であるが、揮発性の記録媒体でも実現可能である。第一話者属性格納部 11 等に、上記の情報が記憶される過程は問わない。例えば、記録媒体を介して上記の情報が第一話者属性格納部 11 等で記憶されるようになってよく、通信回線等を介して送信された上記の情報が第一話者属性格納部 11 等で記憶されるようになってよく、あるいは、入力デバイスを介して入力された上記の情報が第一話者属性格納部 11 等で記憶されるようになってよい。

[0091] 第一話者属性受付部 12、および第二話者属性受付部 22 は、テンキーやキーボード等の入力手段のデバイスドライバーや、メニュー画面の制御ソフトウェア等で実現され得る。

[0092] 第一音声受付部 14、および第二音声受付部 24 は、例えば、マイクとそのデバイスドライバーなどにより実現され得る。

[0093] 第一音声送信部 16、第一音声受信部 17、第一話者属性送信部 19、第二音声送信部 26、第二音声受信部 27、第二話者属性送信部 29、第三話者属性受信部 303、音声情報受信部 306、音声認識結果送信部 310、第三話者属性送信部 311、第四話者属性受信部 403、音声認識結果受信部 406、翻訳結果送信部 410、第四話者属性送信部 411、第五話者属

性受信部 503、翻訳結果受信部 505、および音声合成結果送信部 508 は、通常、無線または有線の通信手段で実現されるが、放送手段または放送受信手段で実現されても良い。

第一音声出力部 18、および第二音声出力部 28 は、スピーカおよびそのドライバーソフト等で実現され得る。

[0094] 第一話者属性蓄積部 13、第一音声認識サーバ選択部 15、第一サーバ選択情報格納手段 151、第二話者属性蓄積部 23、第二音声認識サーバ選択部 25、第二サーバ選択手段 252、音声話者属性取得部 304、第三話者属性蓄積部 305、音声認識モデル選択部 307、音声認識部 308、翻訳サーバ選択部 309、第三モデル選択手段 3072、第三サーバ選択手段 3092、言語話者属性取得部 404、第四話者属性蓄積部 405、翻訳モデル選択部 407、翻訳部 408、音声合成サーバ選択部 409、第四モデル選択手段 4072、第四サーバ選択手段 4092、第五話者属性蓄積部 504、音声合成モデル選択部 506、音声合成部 507、および第五モデル選択手段 5062 は、通常、MPU やメモリ等から実現され得る。第一話者属性蓄積部 13 等の処理手順は、通常、ソフトウェアで実現され、当該ソフトウェアは ROM 等の記録媒体に記録されている。但し、ハードウェア（専用回路）で実現しても良い。

[0095] 次に、音声翻訳システムの動作について図 6 から図 10 のフローチャートを用いて説明する。まず、第一端末装置 1 の動作について図 6 のフローチャートを用いて説明する。

[0096] （ステップ S601）第一話者属性受付部 12 等の受付部は、ユーザ A からの入力を受け付けたか否かを判断する。入力を受け付ければステップ S602 に行き、入力を受け付けなければステップ S606 に行く。

[0097] （ステップ S602）第一話者属性受付部 12 は、ステップ S601 で受け付けた入力の話者属性であるか否かを判断する。話者属性であればステップ S603 に行き、話者属性でなければステップ S604 に行く。

[0098] （ステップ S603）第一話者属性蓄積部 13 は、受け付けられた 1 以上

の話者属性を、第一話者属性格納部 11 に蓄積する。ステップ S 601 に戻る。

[0099] (ステップ S 604) 図示しない受付部は、ステップ S 601 で受け付けた入力が発呼要求であるか否かを判断する。発呼要求であればステップ S 605 に行き、発呼要求でなければステップ S 601 に戻る。なお、発呼要求は、第二端末装置 2 のユーザ B に対しての通話の依頼であり、通常、第二端末装置 2 の第二端末装置識別子（電話番号など）が含まれる。

[0100] (ステップ S 605) 図示しない発呼部は、第二端末装置 2 に対して、発呼する。そして、発呼の結果、通話が開始される。ステップ S 601 に戻る。

[0101] (ステップ S 606) 第一音声受付部 14 は、ユーザ A の音声を受け付けたか否かを判断する。音声を受け付ければステップ S 607 に行き、音声を受け付けなければステップ S 601 に戻る。

[0102] (ステップ S 607) 第一サーバ選択手段 152 は、第一話者属性格納部 11 から、1 以上の話者属性を読み出す。

[0103] (ステップ S 608) 第一サーバ選択手段 152 は、ステップ S 607 で読み出した 1 以上の話者属性を、第一サーバ選択情報格納手段 151 の第一サーバ選択情報（音声認識サーバ選択情報管理表）に適用し、音声認識サーバ装置 3 を選択する。ここで、音声認識サーバ装置 3 の選択とは、例えば、一の音声認識サーバ装置識別子を取得することである。

[0104] (ステップ S 609) 第一話者属性送信部 19 は、第一話者属性格納部 11 に格納されている 1 以上の話者属性を用いて、音声翻訳制御情報を構成する。第一話者属性送信部 19 は、例えば、入力された第二端末装置 2 の電話番号から決定した目的言語の識別子を取得する。また、第一話者属性送信部 19 は、格納されている第一端末装置 1 の電話番号から決定した原言語の識別子を取得する。例えば、電話番号は国コードを含むので、第一話者属性送信部 19 は、かかる国コードから目的言語を決定する。第一話者属性送信部 19 は、国コードと目的言語識別子との対応表（例えば、「81：日本語」

「８２：韓国語」などのレコードを有する表）を保持している。そして、第一話者属性送信部１９は、第一話者属性格納部１１に格納されている１以上の話者属性と、原言語の識別子と目的言語の識別子等から、音声翻訳制御情報を構成する。

[0105] （ステップＳ６１０）第一音声送信部１６は、ステップＳ６０６で受け付けた音声をデジタル化し、音声情報を取得する。そして、第一音声送信部１６は、当該音声情報を、ステップＳ６０８で選択された音声認識サーバ装置３に送信する。

[0106] （ステップＳ６１１）第一話者属性送信部１９は、ステップＳ６０９で構成された音声翻訳制御情報を、ステップＳ６０８で選択された音声認識サーバ装置３に送信する。なお、ここで、第一話者属性送信部１９は、１以上の話者属性のみを、ステップＳ６０８で選択された音声認識サーバ装置３に送信しても良い。ステップＳ６０１に戻る。

[0107] なお、図６のフローチャートにおいて、通話中に、再度、ステップＳ６０７、ステップＳ６０８、ステップＳ６０９、およびステップＳ６１１の処理を行わないことは好適である。つまり、一通話において、一度または、音声情報の送信より少ない回数、ステップＳ６０７、ステップＳ６０８、ステップＳ６０９、およびステップＳ６１１の処理を行うことは好適である。

[0108] さらに、図６のフローチャートにおいて、電源オフや処理終了の割り込みにより処理は終了する。

[0109] また、第二端末装置２の動作は、第一端末装置１の動作と同様であるので、説明を省略する。

[0110] 次に、音声認識サーバ装置３の動作について図７のフローチャートを用いて説明する。

[0111] （ステップＳ７０１）音声情報受信部３０６は、音声情報を受信したか否かを判断する。音声情報を受信すればステップＳ７０２に行き、音声情報を受信しなければステップＳ７０１に戻る。

[0112] （ステップＳ７０２）第三話者属性受信部３０３は、音声翻訳制御情報を

受信したか否かを判断する。音声翻訳制御情報を受信すればステップS 7 0 3に行き、受信しなければステップS 7 1 0に行く。

[0113] (ステップS 7 0 3) 音声話者属性取得部3 0 4は、ステップS 7 0 1で受信された音声情報から、1以上の話者属性を取得する。かかる処理を話者属性取得処理といい、図8のフローチャートを用いて説明する。

[0114] (ステップS 7 0 4) 第三話者属性蓄積部3 0 5は、ステップS 7 0 2で受信した音声翻訳制御情報に、ステップS 7 0 3で取得した1以上の話者属性を加え、新たな音声翻訳制御情報を構成し、当該音声翻訳制御情報を第三話者属性格納部3 0 1に、少なくとも一時的に蓄積する。なお、第三話者属性蓄積部3 0 5は、ステップS 7 0 3で取得したすべての話者属性をステップS 7 0 2で受信した音声翻訳制御情報に加える必要はない。また、第三話者属性蓄積部3 0 5は、ステップS 7 0 3で取得したすべての話者属性を優先し、ステップS 7 0 2で受信した音声翻訳制御情報の一部の話者属性を修正しても良い。ステップS 7 0 3で取得したすべての話者属性を優先させた場合、例えば、女性的な男性の音声に対しても、音声認識しやすくなるなどのメリットがある。

[0115] (ステップS 7 0 5) 第三モデル選択手段3 0 7 2は、第三話者属性格納部3 0 1に格納されている音声翻訳制御情報が有する1以上の話者属性を用いて、音声認識モデル選択情報管理表を検索し、音声認識モデル識別子を取得する。つまり、第三モデル選択手段3 0 7 2は、音声認識モデルを選択する。そして、第三モデル選択手段3 0 7 2は、選択した音声認識モデルを、音声認識モデル格納部3 0 2から読み出す。

[0116] (ステップS 7 0 6) 音声認識部3 0 8は、読み出された音声認識モデルを用いて、ステップS 7 0 1で受信された音声情報に対して、音声認識処理を行う。そして、音声認識部3 0 8は、音声認識結果を得る。

[0117] (ステップS 7 0 7) 第三サーバ選択手段3 0 9 2は、第三話者属性格納部3 0 1に格納されている音声翻訳制御情報が有する1以上の話者属性を用いて、翻訳サーバ選択情報管理表を検索し、1以上の話者属性に対応する翻

訳サーバ装置識別子を取得する。

- [0118] (ステップS 7 0 8) 音声認識結果送信部 3 1 0 は、ステップS 7 0 7 で取得された訳サーバ装置識別子に対応する訳サーバ装置 4 に、ステップS 7 0 6 で得られた音声認識結果を送信する。
- [0119] (ステップS 7 0 9) 第三話者属性送信部 3 1 1 は、第三話者属性格納部 3 0 1 に格納されている音声翻訳制御情報を、ステップS 7 0 7 で取得された訳サーバ装置識別子に対応する訳サーバ装置 4 に送信し、ステップS 7 0 1 に戻る。
- [0120] (ステップS 7 1 0) 第三モデル選択手段 3 0 7 2 は、第三話者属性格納部 3 0 1 に音声翻訳制御情報が格納されているか否かを判断する。音声翻訳制御情報が格納されていればステップS 7 1 1 に行き、格納されていなければステップS 7 1 2 に行く。
- [0121] (ステップS 7 1 1) 第三モデル選択手段 3 0 7 2 は、第三話者属性格納部 3 0 1 に格納されている音声翻訳制御情報を読み出し、ステップS 7 0 5 に行く
- [0122] (ステップS 7 1 2) 第三モデル選択手段 3 0 7 2 は、音声認識モデル格納部 3 0 2 に格納されている任意の音声認識モデルを読み出し、ステップS 7 0 6 に行く。
- [0123] なお、図 7 のフローチャートにおいて、話者属性取得処理で音声認識が行われている場合、再度の音声認識処理を行わなくても良い。ただし、話者属性取得処理で音声認識が行われている場合でも、音声認識モデルを選択し、精度の高い音声認識処理を行うことは好適である。
- [0124] また、図 7 のフローチャートにおいて、ステップS 7 0 3 の話者属性取得処理を、音声認識処理した結果に対して行っても良い。
- [0125] さらに、図 7 のフローチャートにおいて、電源オフや処理終了の割り込みにより処理は終了する。
- [0126] ステップS 7 0 3 の話者属性取得処理を、図 8 のフローチャートを用いて説明する。

- [0127] (ステップS801) 音声話者属性取得部304は、音声情報から1以上の特徴量を取得する(音声分析する)。音声話者属性取得部304が取得した1以上の特徴量から構成されるベクトルである特徴ベクトルデータは、例えば、三角型フィルタを用いたチャンネル数24のフィルタバンク出力を離散コサイン変換したMFCCであり、その静的パラメータ、デルタパラメータおよびデルタデルタパラメータをそれぞれ12次元、さらに正規化されたパワーとデルタパワーおよびデルタデルタパワー(39次元)を有する。
- [0128] (ステップS802) 音声話者属性取得部304は、ステップS801で取得した1以上の特徴量を用いて、話者の性別を決定する。
- [0129] (ステップS803) 音声話者属性取得部304は、ステップS801で取得した1以上の特徴量を用いて、話者の年齢層を決定する。
- [0130] (ステップS804) 音声話者属性取得部304は、音声情報から、話速を取得する。なお、話速を取得する処理は公知技術である。
- [0131] (ステップS805) 音声話者属性取得部304は、音声認識部308に音声認識処理を依頼し、音声認識結果を得る。
- [0132] (ステップS806) 音声話者属性取得部304は、ステップS805で得られた音声認識結果を自然言語処理し、話者クラスを決定する。上位処理にリターンする。
- [0133] なお、図8のフローチャートにおいて、受信された音声翻訳制御情報(話者属性)に含まれている話者属性(例えば、性別)について、取得する処理(例えば、ステップS802)を行わなくても良い。処理の高速化になる。また、会話によって不変の話者属性(性別、年齢層など)は、一度のみ取得し、会話中に変化する話者属性(話速など)は、都度、取得する、または、累積された情報をも用いて、変化させていくことは好適である。
- [0134] 次に、翻訳サーバ装置4の動作について図9のフローチャートを用いて説明する。
- [0135] (ステップS901) 音声認識結果受信部406は、音声認識結果を受信したか否かを判断する。音声認識結果を受信すればステップS902に行き

、音声認識結果を受信しなければステップS 9 0 1に戻る。

[0136] (ステップS 9 0 2) 第四話者属性受信部4 0 3は、音声翻訳制御情報を受信したか否かを判断する。音声翻訳制御情報を受信すればステップS 9 0 3に行き、受信しなければステップS 9 0 9に行く。

[0137] (ステップS 9 0 3) 言語話者属性取得部4 0 4は、ステップS 9 0 1で受信された音声認識結果を自然言語処理し、1以上の言語話者属性を取得する。言語話者属性取得部4 0 4は、例えば、音声認識結果から、話者クラスを取得する。

[0138] (ステップS 9 0 4) 第四話者属性蓄積部4 0 5は、ステップS 9 0 2で受信した音声翻訳制御情報に、ステップS 9 0 3で取得された1以上の言語話者属性を加えて、音声翻訳制御情報を構成し、当該音声翻訳制御情報を、第四話者属性格納部4 0 1に、少なくとも一時的に蓄積する。

[0139] (ステップS 9 0 5) 第四モデル選択手段4 0 7 2は、ステップS 9 0 2で受信した音声翻訳制御情報が有する1以上の話者属性、または第四話者属性格納部4 0 1に格納されている音声翻訳制御情報が有する1以上の話者属性を用いて、翻訳モデル選択情報管理表を検索し、翻訳モデル識別子を取得する。つまり、第四モデル選択手段4 0 7 2は、翻訳モデルを選択する。そして、第四モデル選択手段4 0 7 2は、選択した翻訳モデルを、翻訳モデル格納部4 0 2から読み出す。

[0140] (ステップS 9 0 6) 翻訳部4 0 8は、読み出された翻訳モデルを用いて、ステップS 9 0 1で受信された音声認識結果に対して、翻訳処理を行う。そして、翻訳部4 0 8は、翻訳結果を得る。

[0141] (ステップS 9 0 7) 第四サーバ選択手段4 0 9 2は、第四話者属性格納部4 0 1に格納されている音声翻訳制御情報が有する1以上の話者属性を用いて、音声合成サーバ選択情報管理表を検索し、1以上の話者属性に対応する音声合成サーバ装置識別子を取得する。

[0142] (ステップS 9 0 8) 翻訳結果送信部4 1 0は、ステップS 9 0 7で取得された音声合成サーバ装置識別子に対応する音声合成サーバ装置5に、ステ

ップS 9 0 6 で得られた翻訳結果を送信する。

[0143] (ステップS 9 0 9) 第四話者属性送信部 4 1 1 は、第四話者属性格納部 4 0 1 に格納されている音声翻訳制御情報を、ステップS 9 0 7 で取得された音声合成サーバ装置識別子に対応する音声合成サーバ装置 5 に送信する。ステップS 9 0 1 に戻る。

[0144] (ステップS 9 1 0) 第四モデル選択手段 4 0 7 2 は、第四話者属性格納部 4 0 1 に音声翻訳制御情報が格納されているか否かを判断する。音声翻訳制御情報が格納されていればステップS 9 1 1 に行き、格納されていなければステップS 9 1 2 に行く。

[0145] (ステップS 9 1 1) 第四モデル選択手段 4 0 7 2 は、第四話者属性格納部 4 0 1 に格納されている音声翻訳制御情報を読み出す。ステップS 9 0 5 に行く

[0146] (ステップS 9 1 2) 第四モデル選択手段 4 0 7 2 は、翻訳モデル格納部 4 0 2 に格納されている任意の翻訳モデルを読み出す。ステップS 9 0 6 に行く。

[0147] なお、図 9 のフローチャートにおいて、電源オフや処理終了の割り込みにより処理は終了する。

[0148] 次に、音声合成サーバ装置 5 の動作について図 1 0 のフローチャートを用いて説明する。

[0149] (ステップS 1 0 0 1) 翻訳結果受信部 5 0 5 は、翻訳結果を受信したか否かを判断する。翻訳結果を受信すればステップS 1 0 0 2 に行き、翻訳結果を受信しなければステップS 1 0 0 1 に戻る。

[0150] (ステップS 1 0 0 2) 第五話者属性受信部 5 0 3 は、音声翻訳制御情報を受信したか否かを判断する。音声翻訳制御情報を受信すればステップS 1 0 0 3 に行き、受信しなければステップS 1 0 0 7 に行く。

[0151] (ステップS 1 0 0 3) 第五話者属性蓄積部 5 0 4 は、ステップS 1 0 0 2 で受信した音声翻訳制御情報を、第五話者属性格納部 5 0 1 に、少なくとも一時的に蓄積する。

- [0152] (ステップS 1 0 0 4) 第五モデル選択手段5 0 6 2は、第五話者属性格納部5 0 1に格納されている音声翻訳制御情報が有する1以上の話者属性を用いて、音声合成モデル選択情報管理表を検索し、音声合成モデル識別子を取得する。つまり、第五モデル選択手段5 0 6 2は、音声合成モデルを選択する。そして、第五モデル選択手段5 0 6 2は、選択した音声合成モデルを、音声合成モデル格納部5 0 2から読み出す。
- [0153] (ステップS 1 0 0 5) 音声合成部5 0 7は、読み出された音声合成モデルを用いて、ステップS 1 0 0 1で受信された翻訳結果に対して、音声合成処理を行う。そして、音声合成部5 0 7は、音声合成された音声情報（音声合成結果）を得る。
- [0154] (ステップS 1 0 0 6) 音声合成結果送信部5 0 8は、第二端末装置2に、ステップS 1 0 0 5で得られた音声合成結果を送信する。なお、例えば、第二端末装置2を識別する第二端末装置識別子（例えば、第二端末装置2の電話番号やIPアドレスなど）は、音声翻訳制御情報に含まれる。ステップS 1 0 0 1に戻る。
- [0155] (ステップS 1 0 0 7) 第五モデル選択手段5 0 6 2は、第五話者属性格納部5 0 1に音声翻訳制御情報が格納されているか否かを判断する。音声翻訳制御情報が格納されていればステップS 1 0 0 8に行き、格納されていなければステップS 1 0 0 9に行く。
- [0156] (ステップS 1 0 1 0) 第五モデル選択手段5 0 6 2は、第五話者属性格納部5 0 1に格納されている音声翻訳制御情報を読み出す。ステップS 1 0 0 4に行く
- [0157] (ステップS 1 0 1 1) 第五モデル選択手段5 0 6 2は、音声合成モデル格納部5 0 2に格納されている任意の音声合成モデルを読み出す。ステップS 1 0 0 5に行く。
- [0158] なお、図10のフローチャートにおいて、電源オフや処理終了の割り込みにより処理は終了する。
- [0159] 以下、本実施の形態における音声翻訳システムの具体的な動作について説

明する。音声翻訳システムの概念図は図 1 である。

- [0160] 今、第一端末装置 1 のユーザ A は、日本語を話す、37 歳の女性であり、日本語はネイティブである。また、第二端末装置 2 ユーザ B は、英語を話す、38 歳の男性であり、英語はネイティブである。
- [0161] そして、第一端末装置 1 の第一話者属性格納部 11 には、図 11 に示す第一話者属性管理表が格納されている。第一話者属性管理表は、性別「女性」、年齢「37 歳」、使用言語「日本語」、および Native「yes」が格納されている。
- [0162] また、第二端末装置 2 の第二話者属性格納部 21 には、図 12 に示す第二話者属性管理表が格納されている。
- [0163] また、第一端末装置 1 の第一サーバ選択情報格納手段 151、および第二端末装置 2 の第二サーバ選択情報格納手段 251 には、図 13 に示す音声認識サーバ選択情報管理表が格納されている。音声認識サーバ選択情報管理表は、「ID」「言語」「話者属性」「音声認識サーバ装置識別子」の属性値を有するレコードを 1 以上格納している。「言語」は、音声認識の対象の言語である。「話者属性」は、「性別」「年齢（ここでは、年齢のカテゴリ）」などを有する。「音声認識サーバ装置識別子」は、音声認識サーバ装置 3 と通信するための情報であり、ここでは、IP アドレスである。なお、第一サーバ選択情報格納手段 151 には、音声認識サーバ選択情報管理表のうちの、言語「日本語」に対応するレコードが存在すれば良い。また、第二サーバ選択情報格納手段 251 には、音声認識サーバ選択情報管理表のうちの、言語「英語」に対応するレコードが存在すれば良い。
- [0164] また、音声認識サーバ装置 3 の第三モデル選択情報格納手段 3071 には、図 14 に示す音声認識モデル選択情報管理表が格納されている。音声認識モデル選択情報管理表は、「言語」「話者属性」「音声認識モデル識別子」の属性値を有するレコードを 1 以上格納している。「音声認識モデル識別子」は、音声認識モデルを識別する情報であり、例えば、音声認識モデルを読み出すために利用される。ここでは、例えば、「音声認識モデル識別子」は

、音声認識モデルが格納されているファイル名などである。

- [0165] また、音声認識サーバ装置 3 の第三サーバ選択情報格納手段 3091 には、図 15 に示す翻訳サーバ選択情報管理表が格納されている。翻訳サーバ選択情報管理表は、「ID」「原言語」「目的言語」「話者属性」「翻訳サーバ装置識別子」の属性値を有するレコードを 1 以上格納している。「原言語」は翻訳の元の言語である。「目的言語」は翻訳の先の言語である。「話者属性」は、ここでは、「性別」「年齢」「第一話者クラス」などを有する。「第一話者クラス」は、使う単語の難しさから決定される、話者の知的レベルを推定した情報である。「第一話者クラス」が「A」である場合、話者は知的レベルが高いとする。「第一話者クラス」が「B」または「C」である場合、話者は知的レベルが中程度、または低いとする。知的レベルの推定方法の例は、後述する。「翻訳サーバ装置識別子」は、翻訳サーバ装置 4 と通信するための情報であり、ここでは、IP アドレスである。

- [0166] また、翻訳サーバ装置 4 の第四モデル選択情報格納手段 4071 は、図 16 に示す翻訳モデル選択情報管理表を保持している。翻訳モデル選択情報管理表は、「ID」「原言語」「話者属性」「翻訳モデル識別子」の属性値を有するレコードを 1 以上格納している。「話者属性」は、ここでは、「性別」「年齢」「第二話者クラス」などを有する。「第二話者クラス」は、使用言語について、ネイティブか否かを示す。ネイティブである場合は「Y」、ネイティブでない場合は「N」の属性値となる。「翻訳モデル識別子」は、翻訳モデルを識別する情報であり、例えば、翻訳モデルを読み出すために利用される。ここでは、例えば、「翻訳モデル識別子」は、翻訳モデルが格納されているファイル名などである。

- [0167] また、翻訳サーバ装置 4 の第四サーバ選択情報格納手段 4091 は、図 17 に示す音声合成サーバ選択情報管理表が格納されている。音声合成サーバ選択情報管理表は、「ID」「目的言語」「話者属性」「音声合成サーバ装置識別子」の属性値を有するレコードを 1 以上格納している。「話者属性」は、ここでは、「性別」「年齢」「話速」「第一話者クラス」「第二話者ク

ラス」などを有する。「話速」は、話す速度であり、ここでは、「速い」「中程度」「遅い」の3つのいずれかの値をとり得る。「速い」は、例えば、話すスピードが「5音節／秒」以上である場合、「中程度」は、例えば、話すスピードが「5音節／秒」未満「3音節／秒」以上である場合、「遅い」は、例えば、話すスピードが「3音節／秒」未満である場合である。ただし、「話速」のカテゴリの種類の種類、カテゴリの方法、アルゴリズムなどは問わない。「音声合成サーバ装置識別子」は、音声合成サーバ装置5と通信するための情報であり、ここでは、IPアドレスである。

[0168] さらに、音声合成サーバ装置5の第五モデル選択情報格納手段5061は、図18に示す音声合成モデル選択情報管理表を保持している。音声合成モデル選択情報管理表は、「ID」「目的言語」「話者属性」「音声合成モデル識別子」の属性値を有するレコードを1以上格納している。「話者属性」は、ここでは、「性別」「年齢」「第二話者クラス」などを有する。「話者属性」として、「話速」や「第一話者クラス」を有することはさらに好適である。「音声合成モデル識別子」は、音声合成モデルを識別する情報であり、例えば、音声合成モデルを読み出すために利用される。ここでは、例えば、「音声合成モデル識別子」は、音声合成モデルが格納されているファイル名などである。

[0169] かかる状況において、ユーザAは、ユーザBに電話をかけようとする。そして、ユーザAは、第一端末装置1から、相手（ユーザB）の電話番号などを入力する画面であり、図19の画面を呼び出した。そして、第一端末装置1は、第一話者属性格納部11に格納されている第一話者属性管理表（図11）を読み出し、図19の画面を表示する。そして、ユーザは、相手の使用言語と、相手の電話番号とを入力し、「発信」ボタンを押下した、とする。なお、図19において、自分の電話番号「080-1111-2256」は、図示しない記録媒体に格納されている、とする。

[0170] 次に、第一端末装置1の、図示しない発呼部は、第二端末装置2に対して、発呼する。そして、通話が始まる。

- [0171] 次に、第一端末装置 1 の第一音声受付部 1 4 は、ユーザ A の音声「おはようございます」を受け付ける。
- [0172] 次に、第一サーバ選択手段 1 5 2 は、第一話者属性格納部 1 1 から、図 1 1 の話者属性を読み出す。
- [0173] 次に、第一サーバ選択手段 1 5 2 は、読み出した 1 以上の話者属性「性別：女性，年齢：37 歳，使用言語：日本語・・・」を、図 1 3 の音声認識サーバ選択情報管理表に適用し、「ID=7」のレコードを検索し、音声認識サーバ装置識別子「186. 221. 1. 27」を取得する。
- [0174] 次に、第一話者属性送信部 1 9 は、1 以上の話者属性を用いて、音声翻訳制御情報を構成する。第一話者属性送信部 1 9 は、例えば、図 2 0 に示す音声翻訳制御情報を構成する。この音声翻訳制御情報は、1 以上の話者属性、および図 1 9 の画面からユーザ A により入力された情報（相手の使用言語 [目的言語] ）を含む。さらに、音声翻訳制御情報は、音声認識サーバ装置識別子「186. 221. 1. 27」を含む。
- [0175] 次に、第一音声送信部 1 6 は、受け付けた音声「おはようございます」をデジタル化し、「おはようございます」の音声情報を取得する。そして、第一音声送信部 1 6 は、当該音声情報を、「186. 221. 1. 27」で識別される音声認識サーバ装置 3 に送信する。
- [0176] 次に、第一話者属性送信部 1 9 は、図 2 0 の音声翻訳制御情報を、「186. 221. 1. 27」で識別される音声認識サーバ装置 3 に送信する。
- [0177] 次に、音声認識サーバ装置 3 の音声情報受信部 3 0 6 は、音声情報「おはようございます」を受信する。そして、第三話者属性受信部 3 0 3 は、図 2 0 の音声翻訳制御情報を受信する。
- [0178] 次に、音声話者属性取得部 3 0 4 は、受信された音声情報「おはようございます」から、1 以上の話者属性を取得する。つまり、第三話者属性受信部 3 0 3 は、音声情報「おはようございます」から 1 以上の特徴量を取得する。そして、第三話者属性受信部 3 0 3 は、1 以上の特徴量を用いて、予め決められた情報を取得する。ここで、図 2 0 の音声翻訳制御情報に、性別や年

齢などの話者属性を含むが、第三話者属性受信部 303 は、音声翻訳制御情報と重複する話者属性（性別や年齢など）を取得し、取得した話者属性を優先させて、音声認識や、後の翻訳や、音声合成に利用しても良い。

[0179] 次に、音声話者属性取得部 304 は、音声情報「おはようございます」から、話速を取得する。ここで、音声話者属性取得部 304 は、4 音節／秒であると、判断し、話速「中程度」を取得した、とする。

[0180] 次に、第三モデル選択手段 3072 は、第三話者属性格納部 301 に格納されている音声翻訳制御情報が有する 1 以上の話者属性、および音声話者属性取得部 304 が取得した 1 以上の話者属性（ここでは、話速「中程度」）を用いて、音声認識モデル選択情報管理表（図 14）から、「ID=18」のレコードを検索し、音声認識モデル識別子「JR6」を取得する。そして、第三モデル選択手段 3072 は、選択した音声認識モデル「JR6」を、音声認識モデル格納部 302 から読み出す。

[0181] 音声認識部 308 は、読み出された音声認識モデルを用いて、受信された音声情報に対して、音声認識処理を行い、音声認識結果「おはようございます」を得る。

[0182] 次に、音声話者属性取得部 304 は、音声認識部 308 に音声認識処理を依頼し、音声認識結果「おはようございます」を得る。

[0183] 音声話者属性取得部 304 は、得られた音声認識結果を自然言語処理し、丁寧語であるので、第一話者クラス「A」を取得する。音声話者属性取得部 304 は、例えば、丁寧語を構成する用語「ございます」「です」や、難易度の高い用語「齟齬」「誤謬」などを格納しており、かかる用語の出現頻度、出現割合などにより、第一話者クラスを決定しても良い。音声話者属性取得部 304 は、「おはようございます」を形態素解析し、「おはよう」「ございます」の 2 つの形態素に分割する。そして、音声話者属性取得部 304 は、「ございます」が管理されている用語に合致することを検知する。次に、音声話者属性取得部 304 は、管理用語である割合「50%」と算出する。次に、音声話者属性取得部 304 は、格納している判断条件「A：管理用

語の出現割合が５％以上、Ｂ：管理用語の出現割合が１％以上かつ５％未満、Ｃ：管理用語の出現割合が１％未満」から、第一話者クラス「Ａ」と決定する。なお、次々と会話がなされるごとに、音声話者属性取得部３０４は、管理用語の出現割合を、都度、算出し、第一話者クラスを決定、変更していくことは好適である。

[0184] 次に、第三話者属性蓄積部３０５は、受信した音声翻訳制御情報（図２０）に、話者属性である話速「中程度」、および第一話者クラス「Ａ」を加える。そして、第三話者属性蓄積部３０５は、図２１の音声翻訳制御情報を、第三話者属性格納部３０１に、少なくとも一時的に蓄積する。

[0185] 次に、第三サーバ選択手段３０９２は、第三話者属性格納部３０１に格納されている音声翻訳制御情報（図２１）が有する１以上の話者属性を用いて、翻訳サーバ選択情報管理表（図１５）を検索し、１以上の話者属性に対応する「ＩＤ＝２５」のレコードの翻訳サーバ装置識別子「７７．１２８．５０．８０」を取得する。そして、第三話者属性蓄積部３０５は、翻訳サーバ装置識別子「７７．１２８．５０．８０」を音声翻訳制御情報に加えて、第三話者属性格納部３０１に蓄積する。かかる更新された音声翻訳制御情報を図２２に示す。

[0186] 次に、音声認識結果送信部３１０は、取得された翻訳サーバ装置識別子「７７．１２８．５０．８０」に対応する翻訳サーバ装置４に、音声認識結果「おはようございます」を送信する。

[0187] そして、第三話者属性送信部３１１は、第三話者属性格納部３０１に格納されている音声翻訳制御情報（図２２）を、取得された翻訳サーバ装置識別子「７７．１２８．５０．８０」に対応する翻訳サーバ装置４に送信する。

[0188] 次に、翻訳サーバ装置４の音声認識結果受信部４０６は、音声認識結果「おはようございます」を受信する。そして、第四話者属性受信部４０３は、音声翻訳制御情報（図２２）を受信する。

[0189] 次に、第四話者属性蓄積部４０５は、受信した音声翻訳制御情報（図２２）を、第四話者属性格納部４０１に、少なくとも一時的に蓄積する。

- [0190] 次に、第四モデル選択手段4072は、図22の音声翻訳制御情報が有する1以上の話者属性を用いて、翻訳モデル選択情報管理表（図16）から、「ID=18」のレコードを検索し、翻訳モデル識別子「JT4」を取得する。そして、第四モデル選択手段4072は、「JT4」の翻訳モデルを、翻訳モデル格納部402から読み出す。
- [0191] 次に、翻訳部408は、読み出された翻訳モデル「JT4」を用いて、受信された音声認識結果「おはようございます」に対して、翻訳処理を行う。そして、翻訳部408は、翻訳結果「Good morning.」を得る。
- [0192] 次に、第四サーバ選択手段4092は、図22の音声翻訳制御情報が有する1以上の話者属性を用いて、音声合成サーバ選択情報管理表（図17）から、1以上の話者属性に対応する「ID=33」のレコードを検索し、音声合成サーバ装置識別子「238. 3. 55. 7」を取得する。
- [0193] そして、第四話者属性蓄積部405は、図22の音声翻訳制御情報に、音声合成サーバ装置識別子「238. 3. 55. 7」を加えた音声翻訳制御情報（図23）を構成し、当該音声翻訳制御情報を第四話者属性格納部401に蓄積する。
- [0194] 次に、翻訳結果送信部410は、音声合成サーバ装置識別子「238. 3. 55. 7」に対応する音声合成サーバ装置5に、翻訳結果「Good morning.」を送信する。
- [0195] 次に、第四話者属性送信部411は、図23の音声翻訳制御情報を、音声合成サーバ装置識別子「238. 3. 55. 7」に対応する音声合成サーバ装置5に送信する。
- [0196] 次に、音声合成サーバ装置5の翻訳結果受信部505は、翻訳結果を受信する。また、第五話者属性受信部503は、図23の音声翻訳制御情報を受信する。
- [0197] そして、第五話者属性蓄積部504は、受信した音声翻訳制御情報を、第五話者属性格納部501に、少なくとも一時的に蓄積する。

- [0198] 次に、第五モデル選択手段5062は、第五話者属性格納部501に格納されている音声翻訳制御情報（図23）が有する1以上の話者属性を用いて、音声合成モデル選択情報管理表から、「ID=18」のレコードを検索し、音声合成モデル識別子「JC9」を取得する。そして、第五モデル選択手段5062は、選択した音声合成モデル「JC9」を、音声合成モデル格納部502から読み出す。
- [0199] 次に、音声合成部507は、読み出された音声合成モデルを用いて、翻訳結果「Good morning.」に対して、音声合成処理を行う。そして、音声合成部507は、音声合成された音声情報（音声合成結果）を得る。
- [0200] 次に、音声合成結果送信部508は、第二端末装置2に、得られた音声合成結果を送信する。
- [0201] 次に、第二端末装置2の第二音声受信部27は、音声合成結果「Good morning」を受信する。そして、第二音声出力部28は、音声「Good morning」を出力する。
- [0202] 以上の処理により、ユーザAが発生した「おはようございます」が、第二端末装置2に至るまでに「Good morning」の音声に変換され、第二端末装置2に「Good morning」が出力される。
- [0203] また、第二端末装置2のユーザBが、「Good morning」に答えて「Good morning」と発生した音声は、上記と同様の処理により、「おはよう」に変換され、第一端末装置1に音声「おはよう」が出力される。
- [0204] 以上、本実施の形態によれば、音声認識、翻訳、音声合成の各処理において、話者属性に合致する適切な装置、または適切なモデルを選択できる。その結果、精度の高い、または話者の属性を引き継いだネットワーク型音声翻訳システムが提供できる。
- [0205] また、本実施の形態によれば、会話中に、話速や話者クラスなどの話者属性が変化した場合、当該変化に合わせた装置（音声認識、翻訳、音声合成の

各装置）やモデル（音声認識、翻訳、音声合成の各モデル）が適用され、適切に音声認識、翻訳、音声合成等の処理が行える。

[0206] なお、本実施の形態によれば、音声認識、翻訳、音声合成の各処理において、話者属性に合致する適切な装置、または適切なモデルを、すべて選択できた。しかし、音声認識を行う装置またはモデルのみが選択できても良いし、翻訳を行う装置またはモデルのみが選択できても良いし、音声合成を行う装置またはモデルのみが選択できても良い。

[0207] また、本実施の形態によれば、例えば、ユーザが、翻訳サーバ装置または翻訳モデルを指定した場合は、指定された翻訳サーバ装置または翻訳モデルを使用して、翻訳処理を行うことは好適である。ユーザは、自分で使用したい表現を蓄積した翻訳サーバ装置または翻訳モデルを利用したい場合もあるからである。かかる場合、例えば、第一端末装置 1 に、利用したい翻訳サーバ装置を識別する翻訳サーバ装置識別子または翻訳モデルを識別する翻訳モデル識別子が格納されている。そして、かかる翻訳サーバ装置識別子または翻訳モデル識別子が音声翻訳制御情報に付加される。そして、かかる音声翻訳制御情報は、第一端末装置 1 から、音声認識サーバ装置 3 を経由して、翻訳サーバ装置 4 に送信される。

[0208] また、同様に、本実施の形態において、例えば、ユーザが、音声合成サーバ装置または音声合成モデルを指定した場合は、指定された音声合成サーバ装置または音声合成モデルを使用して、音声合成処理を行うことは好適である。ユーザは、例えば、自分の音声を集めた音声合成モデル、または自分の音声を集めた音声合成モデルを格納した音声合成サーバ装置を利用して、目的言語の音声合成を行いたい場合もあるからである。かかる場合、例えば、第一端末装置 1 に、利用したい音声合成サーバ装置を識別する音声合成サーバ装置識別子または音声合成モデルを識別する音声合成モデル識別子が格納されている。そして、かかる音声合成サーバ装置識別子または音声合成モデル識別子が音声翻訳制御情報に付加される。そして、かかる音声翻訳制御情報は、第一端末装置 1 から、音声認識サーバ装置 3、および翻訳サーバ装置

4 を経由して、音声合成サーバ装置 5 に送信される。

[0209] ユーザが指定した翻訳サーバ装置を識別する情報、およびユーザが指定した音声合成サーバ装置を識別する情報が付加された音声翻訳制御情報の例は、図 24 である。

[0210] また、本実施の形態によれば、第一端末装置 1 が音声認識サーバ装置 3 の選択処理を行った。また、音声認識サーバ装置 3 が、音声認識モデルの選択処理、および翻訳サーバ装置 4 の選択処理を行った。また、翻訳サーバ装置 4 が、翻訳モデルの選択処理、および音声合成サーバ装置 5 の選択処理を行った。さらに、音声合成サーバ装置 5 が、音声合成モデルの選択処理を行った。しかし、かかるモデルやサーバ装置の選択処理は、他の装置が行っても良い。例えば、一の制御装置が、かかるサーバ装置の選択処理を行う場合の、音声翻訳システム 6 の概念図は、図 25 である。図 25 において、音声翻訳システムは、1 以上の第一端末装置 251、1 以上の第二端末装置 252、1 以上の音声認識サーバ装置 253、1 以上の翻訳サーバ装置 254、1 以上の音声合成サーバ装置 5、および制御装置 256 を具備する。この音声翻訳システム 6 と、上述した音声翻訳システムとの差異は、サーバ装置の選択処理を行う装置が異なるために生じる差異である。なお、図 25 において、モデルの選択は、音声認識サーバ装置 253、翻訳サーバ装置 254、および音声合成サーバ装置 5 で、それぞれ行う。

[0211] また、図 25 において、第一端末装置 251、第二端末装置 252、音声認識サーバ装置 253、翻訳サーバ装置 254、および音声合成サーバ装置 5 は、それぞれ、処理前の結果を制御装置 256 から受信し、処理後の結果を制御装置 256 に送信する。つまり、第一端末装置 251 は、ユーザ A から受け付けた音声情報を制御装置 256 に送信する。そして、制御装置 256 は、音声認識を行う音声認識サーバ装置 253 を決定し、音声情報を音声認識サーバ装置 253 に送信する。次に、音声認識サーバ装置 253 は音声情報を受信し、必要に応じて、音声認識モデルを選択し、音声認識処理を行う。そして、音声認識サーバ装置 253 は、音声認識結果を制御装置 256

に送信する。次に、制御装置 256 は、音声認識結果を音声認識サーバ装置 253 から受信し、翻訳を行う翻訳サーバ装置 254 を選択する。そして、制御装置 256 は、選択した翻訳サーバ装置 254 に、音声認識結果を送信する。次に、翻訳サーバ装置 254 は、音声認識結果を受信し、必要に応じて、翻訳モデルを選択し、翻訳処理を行う。そして、翻訳サーバ装置 254 は、翻訳結果を制御装置 256 に送信する。次に、制御装置 256 は、翻訳結果を翻訳サーバ装置 254 から受信し、音声合成を行う音声合成サーバ装置 5 を選択する。そして、制御装置 256 は、選択した音声合成サーバ装置 5 に、翻訳結果を送信する。次に、音声合成サーバ装置 5 は、翻訳結果を受信し、必要に応じて、音声合成モデルを選択し、音声合成処理を行う。そして、音声合成サーバ装置 5 は、音声合成結果を制御装置 256 に送信する。次に、制御装置 256 は、音声合成結果を音声合成サーバ装置 5 から受信し、第二端末装置 252 に送信する。次に、第二端末装置 252 は、音声合成結果を受信し、出力する。

[0212] 図 26 は、音声翻訳システム 6 のブロック図である。図 26 において、第一端末装置 251 の第一音声送信部 16、第一音声受信部 17、および第一話者属性送信部 19、第二端末装置 252 の第二音声送信部 26、第二音声受信部 27、および第二話者属性送信部 29 が情報の送受信を行う装置は、制御装置 256 である。

[0213] 図 26 において、第一端末装置 251 は、第一話者属性格納部 11、第一話者属性受付部 12、第一話者属性蓄積部 13、第一音声受付部 14、第一音声送信部 16、第一音声受信部 17、第一音声出力部 18、第一話者属性送信部 19 を備える。第二端末装置 252 は、第二話者属性格納部 21、第二話者属性受付部 22、第二話者属性蓄積部 23、第二音声受付部 24、第二音声送信部 26、第二音声受信部 27、第二音声出力部 28、第二話者属性送信部 29 を備える。

[0214] 図 27 は、制御装置 256 のブロック図である。制御装置 256 は、話者属性格納部 2561、送受信部 2562、話者属性蓄積部 2563、第二音

声認識サーバ選択部 25、翻訳サーバ選択部 309、音声合成サーバ選択部 409 を備える。話者属性格納部 2561 は、1 以上の話者属性を格納し得る。話者属性格納部 2561 は、音声翻訳制御情報を格納していても良い。送受信部 2562 は、第一端末装置 251、第二端末装置 252、音声認識サーバ装置 253、翻訳サーバ装置 254、および音声合成サーバ装置 5 との間で、各種の情報を送受信する。各種の情報とは、音声情報、音声認識結果、翻訳結果、音声合成結果、および音声翻訳制御情報（一部の話者属性も含む）などである。送受信部 2562 は、通常、無線又は有線の通信手段により実現され得る。話者属性蓄積部 2563 は、送受信部 2562 が受信した 1 以上の話者属性（音声翻訳制御情報でも良い）を、話者属性格納部 2561 に蓄積する。

[0215] また、図 28 は、音声認識サーバ装置 253 のブロック図である。音声認識サーバ装置 253 は、第三話者属性格納部 301、音声認識モデル格納部 302、第三話者属性受信部 303、音声話者属性取得部 304、第三話者属性蓄積部 305、音声情報受信部 306、音声認識モデル選択部 307、音声認識部 308、音声認識結果送信部 310、第三話者属性送信部 311 を備える。

[0216] また、図 29 は、翻訳サーバ装置 254 のブロック図である。翻訳サーバ装置 254 は、第四話者属性格納部 401、翻訳モデル格納部 402、第四話者属性受信部 403、第四話者属性蓄積部 405、音声認識結果受信部 406、翻訳モデル選択部 407、翻訳部 408、翻訳結果送信部 410、第四話者属性送信部 411 を備える。

[0217] また、本実施の形態において、音声翻訳制御情報の例は、図 20 から図 24 を用いて説明した。ただし、音声翻訳制御情報のフォーマットは問わない。音声翻訳制御情報は、図 30 に示すような XML のフォーマットでも良いことは言うまでもない。図 30 に示す音声翻訳制御情報の記述言語を、音声翻訳用マークアップランゲージ STML (Speech Translation Markup Language) という。図 30 において、ユ

ーザID（話者のユーザを特定する情報“Mike”）、音声認識の出力結果のフォーマット、またはサイズ（MaxNBest=“2”）、原言語「英語」（language=“en”）、翻訳対象を特定する情報（ここでは、Task=“Dictation”、会話のドメイン（ここでは、「Travel（旅行）」）、入力音声のフォーマットを示す情報（ここでは、“ADPCM”）が記載されている。また、図30において、話者属性のうちの性別（ここでは、“male”）、年齢（ここでは、“30”）、およびネイティブか否か（ここでは、“no”）が記載されている。また、図30において、出力テキストのフォーマットを示す情報（ここでは、“SurfaceForm”）が記載されている。さらに、音声翻訳制御情報において、出力音声のフォーマットを示す情報、入出力音声の声質を指定する情報、入力テキストのフォーマットを示す情報などが存在しても良い。また、上記の（MaxNBest=“2”）は、音声認識結果の上位2番目までの候補を出力し、送信することを示している。なお、Nbestは音声認識結果の上位N番目までの候補という意味である。

- [0218] また、本実施の形態において、音声認識サーバ装置3、および翻訳サーバ装置4は、それぞれ、音声認識モデルの選択や翻訳モデルの選択を行わなくても良い。また、音声認識処理を行う音声認識サーバ装置3や、翻訳処理を行う翻訳サーバ装置4の選択も行わなくても良い。かかる場合、話者属性に応じた音声合成サーバ装置5の選択や、音声合成モデルの選択が行われる。かかる音声合成サーバ装置5の選択処理、音声合成モデルの選択処理は、上述した通りである。また、音声合成サーバ装置5の音声合成部507は、話者属性（例えば、話速や音の高低や音質など）に応じて、音声を変換して、出力する音声情報を構成しても良い。つまり、音声合成部507は、翻訳結果受信部505が受信した翻訳結果を、第五話者属性受信部503が受信した1以上の話者属性が示す属性に合致するように、音声合成モデル格納部502の音声合成モデルを用いて音声合成し、音声合成結果を取得しても良い。また、音声合成部507は、翻訳結果受信部505が受信した翻訳結果を

、音声翻訳制御情報の話者属性が示す属性に合致するように、音声合成モデル格納部52の音声合成モデルを用いて音声合成し、音声合成結果を取得しても良い。かかる場合も、音声合成モデルの選択と言っても良い。

[0219] さらに、本実施の形態における処理は、ソフトウェアで実現しても良い。そして、このソフトウェアをソフトウェアダウンロード等により配布しても良い。また、このソフトウェアをCD-ROMなどの記録媒体に記録して流布しても良い。なお、このことは、本明細書における他の実施の形態においても該当する。なお、本実施の形態における第一端末装置を実現するソフトウェアは、以下のようなプログラムである。つまり、このプログラムは、コンピュータを、音声を受け付ける第一音声受付部と、記憶媒体に格納している1以上の話者属性に応じて、2以上の音声認識サーバ装置のうちの音声認識サーバ装置を選択する第一音声認識サーバ選択部と、前記第一音声認識サーバ選択部が選択した音声認識サーバ装置に、前記第一音声受付部が受け付けた音声から構成される音声情報を送信する第一音声送信部として機能させるためのプログラム、である。

[0220] また、本実施の形態における音声認識サーバ装置を実現するソフトウェアは、コンピュータを、音声情報を受信する音声情報受信部と、記憶媒体に格納している1以上の話者属性に応じて、記憶媒体に格納している2以上の音声認識モデルから、一の音声認識モデルを選択する音声認識モデル選択部と、前記音声情報受信部が受信した音声情報を、前記音声認識モデル選択部が選択した音声認識モデルを用いて音声認識し、音声認識結果を取得する音声認識部と、前記音声認識結果を送信する音声認識結果送信部として機能させるためのプログラム、である。

[0221] また、本実施の形態における音声認識サーバ装置を実現するソフトウェアは、コンピュータを、音声情報を受信する音声情報受信部と、前記音声情報受信部が受信した音声情報を、記憶媒体に格納している音声認識モデルを用いて音声認識し、音声認識結果を取得する音声認識部と、記憶媒体に格納している1以上の話者属性に応じて、2以上の翻訳サーバ装置のうちの翻訳

サーバ装置を選択する翻訳サーバ選択部と、前記翻訳サーバ選択部が選択した翻訳サーバ装置に、前記音声認識結果を送信する音声認識結果送信部として機能させるためのプログラム、である。

[0222] また、本実施の形態における音声認識サーバ装置を実現するソフトウェアは、コンピュータを、前記音声情報受信部が受信した音声情報から、1以上の音声に関する話者属性を取得する音声話者属性取得部と、前記音声話者属性取得部が取得した1以上の話者属性を、記憶媒体に蓄積する第三話者属性蓄積部として、さらに機能させるためのプログラム、である。

[0223] また、本実施の形態における翻訳サーバ装置を実現するソフトウェアは、コンピュータを、1以上の話者属性を受信する第四話者属性受信部と、音声認識結果を受信する音声認識結果受信部と、前記第四話者属性受信部が受信した1以上の話者属性に応じて、記憶媒体に格納している2以上の翻訳モデルから、一の翻訳モデルを選択する翻訳モデル選択部と、前記音声認識結果受信部が受信した音声認識結果を、前記翻訳モデル選択部が選択した翻訳モデルを用いて、目的言語に翻訳し、翻訳結果を取得する翻訳部と、前記翻訳結果を送信する翻訳結果送信部として機能させるためのプログラム、である。

[0224] また、本実施の形態における翻訳サーバ装置を実現するソフトウェアは、コンピュータを、1以上の話者属性を受信する第四話者属性受信部と、音声認識結果を受信する音声認識結果受信部と、前記音声認識結果受信部が受信した音声認識結果を、記憶媒体に格納している翻訳モデルを用いて、目的言語に翻訳し、翻訳結果を取得する翻訳部と、前記1以上の話者属性に応じて、2以上の音声合成サーバ装置のうちの音声合成サーバ装置を選択する音声合成サーバ選択部と、前記音声合成サーバ選択部が選択した音声合成サーバ装置に、前記翻訳結果を送信する翻訳結果送信部として機能させるためのプログラム、である。

[0225] また、本実施の形態における翻訳サーバ装置を実現するソフトウェアは、コンピュータを、前記音声認識結果受信部が受信した音声認識結果から、1

以上の言語に関する話者属性を取得する言語話者属性取得部と、前記言語話者属性取得部が取得した１以上の話者属性を、記憶媒体に蓄積する第四話者属性蓄積部として機能させるためのプログラム、である。

[0226] また、本実施の形態における音声合成サーバ装置を実現するソフトウェアは、コンピュータを、１以上の話者属性を受信する第五話者属性受信部と、翻訳結果を受信する翻訳結果受信部と、前記第五話者属性受信部が受信した１以上の話者属性に応じて、記憶媒体に格納している２以上の音声合成モデルから、一の音声合成モデルを選択する音声合成モデル選択部と、前記翻訳結果受信部が受信した翻訳結果を、前記音声合成モデル選択部が選択した音声合成モデルを用いて音声合成し、音声合成結果を取得する音声合成部と、前記音声合成結果を第二端末装置に送信する音声合成結果送信部として機能させるためのプログラム、である。

[0227] また、図３１は、本明細書で述べたプログラムを実行して、上述した実施の形態の音声翻訳システム等を実現するコンピュータの外観を示す。上述の実施の形態は、コンピュータハードウェア及びその上で実行されるコンピュータプログラムで実現され得る。図３１は、このコンピュータシステム３４０の概観図であり、図３２は、コンピュータシステム３４０の内部構成を示す図である。

[0228] 図３１において、コンピュータシステム３４０は、ＦＤドライブ３４１１、ＣＤ－ＲＯＭドライブ３４１２を含むコンピュータ３４１と、キーボード３４２と、マウス３４３と、モニタ３４４とを含む。

[0229] 図３２において、コンピュータ３４１は、ＦＤドライブ３４１１、ＣＤ－ＲＯＭドライブ３４１２に加えて、ＭＰＵ３４１３と、ＣＤ－ＲＯＭドライブ３４１２及びＦＤドライブ３４１１に接続されたバス３４１４と、ブートアッププログラム等のプログラムを記憶するためのＲＯＭ３４１５に接続され、アプリケーションプログラムの命令を一時的に記憶するとともに一時記憶空間を提供するためのＲＡＭ３４１６と、アプリケーションプログラム、システムプログラム、及びデータを記憶するためのハードディスク３４１７

とを含む。ここでは、図示しないが、コンピュータ 341 は、さらに、LAN への接続を提供するネットワークカードを含んでも良い。

[0230] コンピュータシステム 340 に、上述した実施の形態の音声翻訳システム等の機能を実行させるプログラムは、CD-ROM 3501、または FD 3502 に記憶されて、CD-ROM ドライブ 3412 または FD ドライブ 3411 に挿入され、さらにハードディスク 3417 に転送されても良い。これに代えて、プログラムは、図示しないネットワークを介してコンピュータ 341 に送信され、ハードディスク 3417 に記憶されても良い。プログラムは実行の際に RAM 3416 にロードされる。プログラムは、CD-ROM 3501、FD 3502 またはネットワークから直接、ロードされても良い。

[0231] プログラムは、コンピュータ 341 に、上述した実施の形態の音声翻訳システム等の機能を実行させるオペレーティングシステム（OS）、またはサードパーティープログラム等は、必ずしも含まなくても良い。プログラムは、制御された態様で適切な機能（モジュール）を呼び出し、所望の結果が得られるようにする命令の部分のみを含んでいれば良い。コンピュータシステム 340 がどのように動作するかは周知であり、詳細な説明は省略する。

[0232] なお、上記プログラムにおいて、情報を送信するステップや、情報をする受信ステップなどでは、ハードウェアによって行われる処理、例えば、送信するステップにおけるモデムやインターフェースカードなどで行われる処理（ハードウェアでしか行われない処理）は含まれない。

[0233] また、上記プログラムを実行するコンピュータは、単数であってもよく、複数であってもよい。すなわち、集中処理を行ってもよく、あるいは分散処理を行ってもよい。

[0234] また、上記各実施の形態において、一の装置に存在する 2 以上の通信手段（音声認識結果受信部、第四話者属性受信部など）は、物理的に一の媒体で実現されても良いことは言うまでもない。

[0235] また、上記各実施の形態において、各処理（各機能）は、単一の装置（シ

ステム)によって集中処理されることによって実現されてもよく、あるいは、複数の装置によって分散処理されることによって実現されてもよい。なお、各処理(各機能)が単一の装置(システム)によって集中処理される場合、音声翻訳システムは、一つの装置であり、音声認識サーバ装置、翻訳サーバ装置、および音声合成サーバ装置は、一つの装置内に含まれる。かかる場合、上記の情報の送信および受信は、情報の受け渡しになる。つまり、上記の受信または送信は広く解する。

さらに具体的には、音声翻訳システムが単一の装置によって集中処理される場合、当該音声翻訳システムは、例えば、図33に示す構成になる。

つまり、音声翻訳システムは、音声受付部3301、第三話者属性格納部301、音声認識モデル格納部302、音声話者属性取得部304、音声認識モデル選択部307、音声認識部308、翻訳モデル格納部402、言語話者属性取得部404、翻訳モデル選択部407、翻訳部408、音声合成モデル格納部502、音声合成モデル選択部506、音声合成部507、音声合成結果出力部3302を備える。

音声受付部3301は、ユーザから音声を受け付ける。この音声は、音声翻訳対象の音声である。音声受付部3301は、例えば、マイクとそのドライバソフトウェア等から構成され得る。

第三話者属性格納部301は、ここでは、通常、ユーザから受け付けた話者属性を格納している。ここでの話者属性は、通常、静的話者属性情報である。

音声話者属性取得部304は、音声受付部3301が受け付けた音声から構成された音声情報から、1以上の音声話者属性を取得する。ここで取得する音声話者属性は、主として、動的話者属性情報であるが、静的話者属性情報でも良い。

音声認識モデル選択部307は、第三話者属性格納部301の話者属性または音声話者属性取得部304が取得した話者属性のうちの、1以上の話者属性に応じて、2以上の音声認識モデルから、一の音声認識モデルを選択す

る。

音声認識部 308 は、音声受付部 3301 が受け付けた音声から構成された音声情報を、音声認識モデル格納部 302 の音声認識モデルを用いて音声認識し、音声認識結果を取得する。また、音声認識部 308 は、音声情報を、音声認識モデル選択部 307 が選択した音声認識モデルを用いて音声認識し、音声認識結果を取得することは好適である。

言語話者属性取得部 404 は、音声認識部 308 が取得した音声認識結果から 1 以上の言語話者属性を取得する。

翻訳モデル選択部 407 は、1 以上の話者属性に応じて、2 以上の翻訳モデルから、一の翻訳モデルを選択する。ここでの話者属性は、第三話者属性格納部 301 の話者属性または音声話者属性取得部 304 が取得した話者属性または言語話者属性取得部 404 が取得した言語話者属性のうちの、1 以上の話者属性である。

翻訳部 408 は、音声認識結果を、翻訳モデル格納部 402 の翻訳モデルを用いて、目的言語に翻訳し、翻訳結果を取得する。翻訳部 408 は、音声認識結果を、翻訳モデル選択部 407 が選択した翻訳モデルを用いて、目的言語に翻訳し、翻訳結果を取得することは好適である。

音声合成モデル選択部 506 は、1 以上の話者属性に応じて、2 以上の音声合成モデルから、一の音声合成モデルを選択する。ここでの話者属性は、第三話者属性格納部 301 の話者属性または音声話者属性取得部 304 が取得した話者属性または言語話者属性取得部 404 が取得した言語話者属性のうちの、1 以上の話者属性である。

音声合成部 507 は、翻訳結果を、音声合成モデル格納部 502 の音声合成モデルを用いて音声合成し、音声合成結果を取得する。音声合成部 507 は、翻訳結果を、音声合成モデル選択部 506 が選択した音声合成モデルを用いて音声合成し、音声合成結果を取得することは好適である。

音声合成結果出力部 3302 は、音声合成部 507 が取得した音声合成結果を出力する。ここでの出力とは、スピーカー等を用いた音声出力、外部の

装置（通常、音声出力装置）への送信、記録媒体への蓄積、他の処理装置や他のプログラムなどへの処理結果の引渡しなどを含む概念である。音声合成結果出力部 3302 は、スピーカーとそのドライバーソフトウェア等から構成され得る。

なお、音声翻訳システムは、第三話者属性格納部 301、音声話者属性取得部 304、音声認識モデル選択部 307、言語話者属性取得部 404、翻訳モデル選択部 407、音声合成モデル選択部 506 は、必須の構成要素ではない。

[0236] 本発明は、以上の実施の形態に限定されることなく、種々の変更が可能であり、それらも本発明の範囲内に包含されるものであることは言うまでもない。

産業上の利用可能性

[0237] 以上のように、本発明にかかる音声翻訳システムは、音声翻訳において、話者の属性に応じて音声認識や翻訳や音声合成を行う装置やモデルを変更でき、音声認識や翻訳や音声合成の各処理において、精度が向上し、適切な出力が行われる、という効果を有し、音声翻訳システム等として有用である。

図面の簡単な説明

[0238] [図1]実施の形態 1 における音声翻訳システムの概念図

[図2] 同音声翻訳システムのブロック図

[図3] 同音声認識サーバ装置のブロック図

[図4] 同翻訳サーバ装置のブロック図

[図5] 同音声合成サーバ装置のブロック図

[図6] 同音声翻訳システムの動作について説明するフローチャート

[図7] 同音声認識サーバ装置の動作について説明するフローチャート

[図8] 同話者属性取得処理の動作について説明するフローチャート

[図9] 同翻訳サーバ装置の動作について説明するフローチャート

[図10] 同音声翻訳システムの動作について説明するフローチャート

[図11] 同第一話者属性管理表を示す図

- [図12] 同第二話者属性管理表を示す図
- [図13] 同音声認識サーバ選択情報管理表を示す図
- [図14] 同音声認識モデル選択情報管理表を示す図
- [図15] 同翻訳サーバ選択情報管理表を示す図
- [図16] 同翻訳モデル選択情報管理表を示す図
- [図17] 同音声合成サーバ選択情報管理表を示す図
- [図18] 同音声合成モデル選択情報管理表を示す図
- [図19] 同第一端末装置の話者属性等の入力画面を示す図
- [図20] 同音声翻訳制御情報の例を示す図
- [図21] 同更新された音声翻訳制御情報の例を示す図
- [図22] 同更新された音声翻訳制御情報の例を示す図
- [図23] 同更新された音声翻訳制御情報の例を示す図
- [図24] 同更新された音声翻訳制御情報の例を示す図
- [図25] 同他の音声翻訳システムの概念図
- [図26] 同他の音声翻訳システムのブロック図
- [図27] 同制御装置のブロック図
- [図28] 同音声認識サーバ装置のブロック図
- [図29] 同翻訳サーバ装置のブロック図
- [図30] 同音声翻訳制御情報の例であるS T M Lの例を示す図
- [図31] 同コンピュータシステムの概観図
- [図32] 同コンピュータシステムのブロック図
- [図33] 同音声翻訳システムの他のブロック図

請求の範囲

[請求項1]

音声を入力する第一端末装置、2以上の音声認識サーバ装置、1以上の翻訳サーバ装置、1以上の音声合成サーバ装置とを有する音声翻訳システムであって、

前記第一端末装置は、

話者の属性値である話者属性を、1以上格納し得る第一話者属性格納部と、

音声を受け付ける第一音声受付部と、

前記1以上の話者属性に応じて、前記2以上の音声認識サーバ装置のうち一の音声認識サーバ装置を選択する第一音声認識サーバ選択部と、

前記第一音声認識サーバ選択部が選択した音声認識サーバ装置に、前記第一音声受付部が受け付けた音声から構成される音声情報を送信する第一音声送信部とを具備し、

前記音声認識サーバ装置は、

前記2以上の言語のうちのすべての言語または2以上の一部の言語について、音声認識モデルを格納し得る音声認識モデル格納部と、

前記音声情報を受信する音声情報受信部と、

前記音声情報受信部が受信した音声情報を、前記音声認識モデル格納部の音声認識モデルを用いて音声認識し、音声認識結果を取得する音声認識部と、

前記音声認識結果を送信する音声認識結果送信部とを具備し、

前記翻訳サーバ装置は、

前記2以上の言語のうちのすべての言語または2以上の一部の言語について、翻訳モデルを格納し得る翻訳モデル格納部と、

前記音声認識結果を受信する音声認識結果受信部と、

前記音声認識結果受信部が受信した音声認識結果を、前記翻訳モデル格納部の翻訳モデルを用いて、目的言語に翻訳し、翻訳結果を取得す

る翻訳部と、
前記翻訳結果を送信する翻訳結果送信部とを具備し、
前記音声合成サーバ装置は、
前記２以上の言語のうちのすべての言語または２以上の一部の言語について、音声合成モデルを格納し得る音声合成モデル格納部と、
前記翻訳結果を受信する翻訳結果受信部と、
前記翻訳結果受信部が受信した翻訳結果を、前記音声合成モデル格納部の音声合成モデルを用いて音声合成し、音声合成結果を取得する音声合成部と、
前記音声合成結果を第二端末装置に送信する音声合成結果送信部とを具備する音声翻訳システム。

[請求項2]

音声を入力する第一端末装置、１以上の音声認識サーバ装置、１以上の翻訳サーバ装置、１以上の音声合成サーバ装置とを有する音声翻訳システムであって、
前記第一端末装置は、
音声を受け付ける第一音声受付部と、
前記第一音声受付部が受け付けた音声から構成される音声情報を、前記音声認識サーバ装置に送信する第一音声送信部とを具備し、
前記音声認識サーバ装置は、
話者の属性値である話者属性を、１以上格納し得る第三話者属性格納部と、
前記２以上の言語のうちのすべての言語または２以上の一部の言語について、２以上の音声認識モデルを格納し得る音声認識モデル格納部と、
前記音声情報を受信する音声情報受信部と、
前記１以上の話者属性に応じて、前記２以上の音声認識モデルから、一の音声認識モデルを選択する音声認識モデル選択部と、
前記音声情報受信部が受信した音声情報を、前記音声認識モデル選択

部が選択した音声認識モデルを用いて音声認識し、音声認識結果を取得する音声認識部と、
前記音声認識結果を送信する音声認識結果送信部とを具備し、
前記翻訳サーバ装置は、
前記 2 以上の言語のうちのすべての言語または 2 以上の一部の言語について、翻訳モデルを格納し得る翻訳モデル格納部と、
前記音声認識結果を受信する音声認識結果受信部と、
前記音声認識結果受信部が受信した音声認識結果を、前記翻訳モデル格納部の翻訳モデルを用いて、目的言語に翻訳し、翻訳結果を取得する翻訳部と、
前記翻訳結果を送信する翻訳結果送信部とを具備し、
前記音声合成サーバ装置は、
前記 2 以上の言語のうちのすべての言語または 2 以上の一部の言語について、音声合成モデルを格納し得る音声合成モデル格納部と、
前記翻訳結果を受信する翻訳結果受信部と、
前記翻訳結果受信部が受信した翻訳結果を、前記音声合成モデル格納部の音声合成モデルを用いて音声合成し、音声合成結果を取得する音声合成部と、
前記音声合成結果を第二端末装置に送信する音声合成結果送信部とを具備する音声翻訳システム。

[請求項3]

1 以上の音声認識サーバ装置、2 以上の翻訳サーバ装置、1 以上の音声合成サーバ装置とを有する音声翻訳システムであって、
前記音声認識サーバ装置は、
話者の属性値である話者属性を、1 以上格納し得る第三話者属性格納部と、
前記 2 以上の言語のうちのすべての言語または 2 以上の一部の言語について、音声認識モデルを格納し得る音声認識モデル格納部と、
音声情報を受信する音声情報受信部と、

前記音声情報受信部が受信した音声情報を、前記音声認識モデル格納部の音声認識モデルを用いて音声認識し、音声認識結果を取得する音声認識部と、

前記 1 以上の話者属性に応じて、前記 2 以上の翻訳サーバ装置のうちの翻訳サーバ装置を選択する翻訳サーバ選択部と、

前記翻訳サーバ選択部が選択した翻訳サーバ装置に、前記音声認識結果を送信する音声認識結果送信部とを具備し、

前記翻訳サーバ装置は、

前記 2 以上の言語のうちのすべての言語または 2 以上の一部の言語について、翻訳モデルを格納し得る翻訳モデル格納部と、

前記音声認識結果を受信する音声認識結果受信部と、

前記音声認識結果受信部が受信した音声認識結果を、前記翻訳モデル格納部の翻訳モデルを用いて、目的言語に翻訳し、翻訳結果を取得する翻訳部と、

前記翻訳結果を送信する翻訳結果送信部とを具備し、

前記音声合成サーバ装置は、

前記 2 以上の言語のうちのすべての言語または 2 以上の一部の言語について、音声合成モデルを格納し得る音声合成モデル格納部と、

前記翻訳結果を受信する翻訳結果受信部と、

前記翻訳結果受信部が受信した翻訳結果を、前記音声合成モデル格納部の音声合成モデルを用いて音声合成し、音声合成結果を取得する音声合成部と、

前記音声合成結果を第二端末装置に送信する音声合成結果送信部とを具備する音声翻訳システム。

[請求項4]

1 以上の音声認識サーバ装置、1 以上の翻訳サーバ装置、1 以上の音声合成サーバ装置とを有する音声翻訳システムであって、

前記音声認識サーバ装置は、

前記 2 以上の言語のうちのすべての言語または 2 以上の一部の言語に

ついて、音声認識モデルを格納し得る音声認識モデル格納部と、
音声情報を受信する音声情報受信部と、
前記音声情報受信部が受信した音声情報を、前記音声認識モデル格納部の音声認識モデルを用いて音声認識し、音声認識結果を取得する音声認識部と、
前記翻訳サーバ装置に、前記音声認識結果を送信する音声認識結果送信部とを具備し、
前記翻訳サーバ装置は、
前記 2 以上の言語のうちのすべての言語または 2 以上の一部の言語について、2 以上の翻訳モデルを格納し得る翻訳モデル格納部と、
1 以上の話者属性を格納し得る第四話者属性格納部と、
前記音声認識結果を受信する音声認識結果受信部と、
前記 1 以上の話者属性に応じて、前記 2 以上の翻訳モデルから、一の翻訳モデルを選択する翻訳モデル選択部と、
前記音声認識結果受信部が受信した音声認識結果を、前記翻訳モデル選択部が選択した翻訳モデルを用いて、目的言語に翻訳し、翻訳結果を取得する翻訳部と、
前記翻訳結果を送信する翻訳結果送信部とを具備し、
前記音声合成サーバ装置は、
前記 2 以上の言語のうちのすべての言語または 2 以上の一部の言語について、音声合成モデルを格納し得る音声合成モデル格納部と、
前記翻訳結果を受信する翻訳結果受信部と、
前記翻訳結果受信部が受信した翻訳結果を、前記音声合成モデル格納部の音声合成モデルを用いて音声合成し、音声合成結果を取得する音声合成部と、
前記音声合成結果を第二端末装置に送信する音声合成結果送信部とを具備する音声翻訳システム。

[請求項5]

1 以上の音声認識サーバ装置、1 以上の翻訳サーバ装置、2 以上の音

声合成サーバ装置とを有する音声翻訳システムであって、
前記音声認識サーバ装置は、
前記 2 以上の言語のうちのすべての言語または 2 以上の一部の言語について、音声認識モデルを格納し得る音声認識モデル格納部と、
音声情報を受信する音声情報受信部と、
前記音声情報受信部が受信した音声情報を、前記音声認識モデル格納部の音声認識モデルを用いて音声認識し、音声認識結果を取得する音声認識部と、
前記翻訳サーバ装置に、前記音声認識結果を送信する音声認識結果送信部とを具備し、
前記翻訳サーバ装置は、
前記 2 以上の言語のうちのすべての言語または 2 以上の一部の言語について、翻訳モデルを格納し得る翻訳モデル格納部と、
1 以上の話者属性を格納し得る第四話者属性格納部と、
前記音声認識結果を受信する音声認識結果受信部と、
前記音声認識結果受信部が受信した音声認識結果を、前記翻訳モデル格納部の翻訳モデルを用いて、目的言語に翻訳し、翻訳結果を取得する翻訳部と、
前記 1 以上の話者属性に応じて、前記 2 以上の音声合成サーバ装置のうち一の音声合成サーバ装置を選択する音声合成サーバ選択部と、
前記音声合成サーバ選択部が選択した音声合成サーバ装置に、前記翻訳結果を送信する翻訳結果送信部とを具備し、
前記音声合成サーバ装置は、
前記 2 以上の言語のうちのすべての言語または 2 以上の一部の言語について、音声合成モデルを格納し得る音声合成モデル格納部と、
前記翻訳結果を受信する翻訳結果受信部と、
前記翻訳結果受信部が受信した翻訳結果を、前記音声合成モデル格納部の音声合成モデルを用いて音声合成し、音声合成結果を取得する音

声合成部と、

前記音声合成結果を第二端末装置に送信する音声合成結果送信部とを具備する音声翻訳システム。

[請求項6]

1以上の音声認識サーバ装置、1以上の翻訳サーバ装置、1以上の音声合成サーバ装置とを有する音声翻訳システムであって、

前記音声認識サーバ装置は、

前記2以上の言語のうちのすべての言語または2以上の一部の言語について、音声認識モデルを格納し得る音声認識モデル格納部と、

音声情報を受信する音声情報受信部と、

前記音声情報受信部が受信した音声情報を、前記音声認識モデル格納部の音声認識モデルを用いて音声認識し、音声認識結果を取得する音声認識部と、

前記翻訳サーバ装置に、前記音声認識結果を送信する音声認識結果送信部とを具備し、

前記翻訳サーバ装置は、

前記2以上の言語のうちのすべての言語または2以上の一部の言語について、翻訳モデルを格納し得る翻訳モデル格納部と、

前記音声認識結果を受信する音声認識結果受信部と、

前記音声認識結果受信部が受信した音声認識結果を、前記翻訳モデル格納部の翻訳モデルを用いて、目的言語に翻訳し、翻訳結果を取得する翻訳部と、

前記音声合成サーバ装置に、前記翻訳結果を送信する翻訳結果送信部とを具備し、

前記音声合成サーバ装置は、

前記2以上の言語のうちのすべての言語または2以上の一部の言語について、2以上の音声合成モデルを格納し得る音声合成モデル格納部と、

1以上の話者属性を格納し得る第五話者属性格納部と、

前記翻訳結果を受信する翻訳結果受信部と、
前記 1 以上の話者属性に応じて、前記 2 以上の音声合成モデルから、
一の音声合成モデルを選択する音声合成モデル選択部と、
前記翻訳結果受信部が受信した翻訳結果を、前記音声合成モデル選択部が選択した音声合成モデルを用いて音声合成し、音声合成結果を取得する音声合成部と、
前記音声合成結果を第二端末装置に送信する音声合成結果送信部とを
具備する音声翻訳システム。

[請求項7] 前記第一端末装置は、
1 以上の話者属性を受け付ける第一話者属性受付部と、
前記 1 以上の話者属性を、前記第一話者属性格納部に蓄積する第一話者属性蓄積部とを具備する請求項 1 記載の音声翻訳システム。

[請求項8] 前記音声認識サーバ装置は、
前記音声情報受信部が受信した音声情報から、1 以上の音声に関する話者属性を取得する音声話者属性取得部と、
前記音声話者属性取得部が取得した 1 以上の話者属性を、第三話者属性格納部に蓄積する第三話者属性蓄積部とをさらに具備する請求項 2 または請求項 3 記載の音声翻訳システム。

[請求項9] 前記翻訳サーバ装置は、
前記音声認識結果受信部が受信した音声認識結果から、1 以上の言語に関する話者属性を取得する言語話者属性取得部と、
前記言語話者属性取得部が取得した 1 以上の話者属性を前記第四話者属性格納部に蓄積する第四話者属性蓄積部とをさらに具備する請求項 4 または請求項 5 記載の音声翻訳システム。

[請求項10] 前記話者が使用する言語である原言語を特定する原言語識別子、および翻訳先の言語である目的言語を特定する目的言語識別子、および 1 以上の話者属性を含む音声翻訳制御情報が、前記音声認識サーバ装置から前記 1 以上の翻訳サーバ装置を経由して、前記音声合成サーバ装

置に送信され、

前記音声認識サーバ選択部、または前記音声認識部、または音声認識モデル選択部、または前記翻訳サーバ選択部、または前記翻訳部、または翻訳モデル選択部、前記音声合成サーバ選択部、または前記音声合成部、または音声合成モデル選択部は、

前記音声翻訳制御情報を用いて、各々の処理を行う請求項 1 記載の音声翻訳システム。

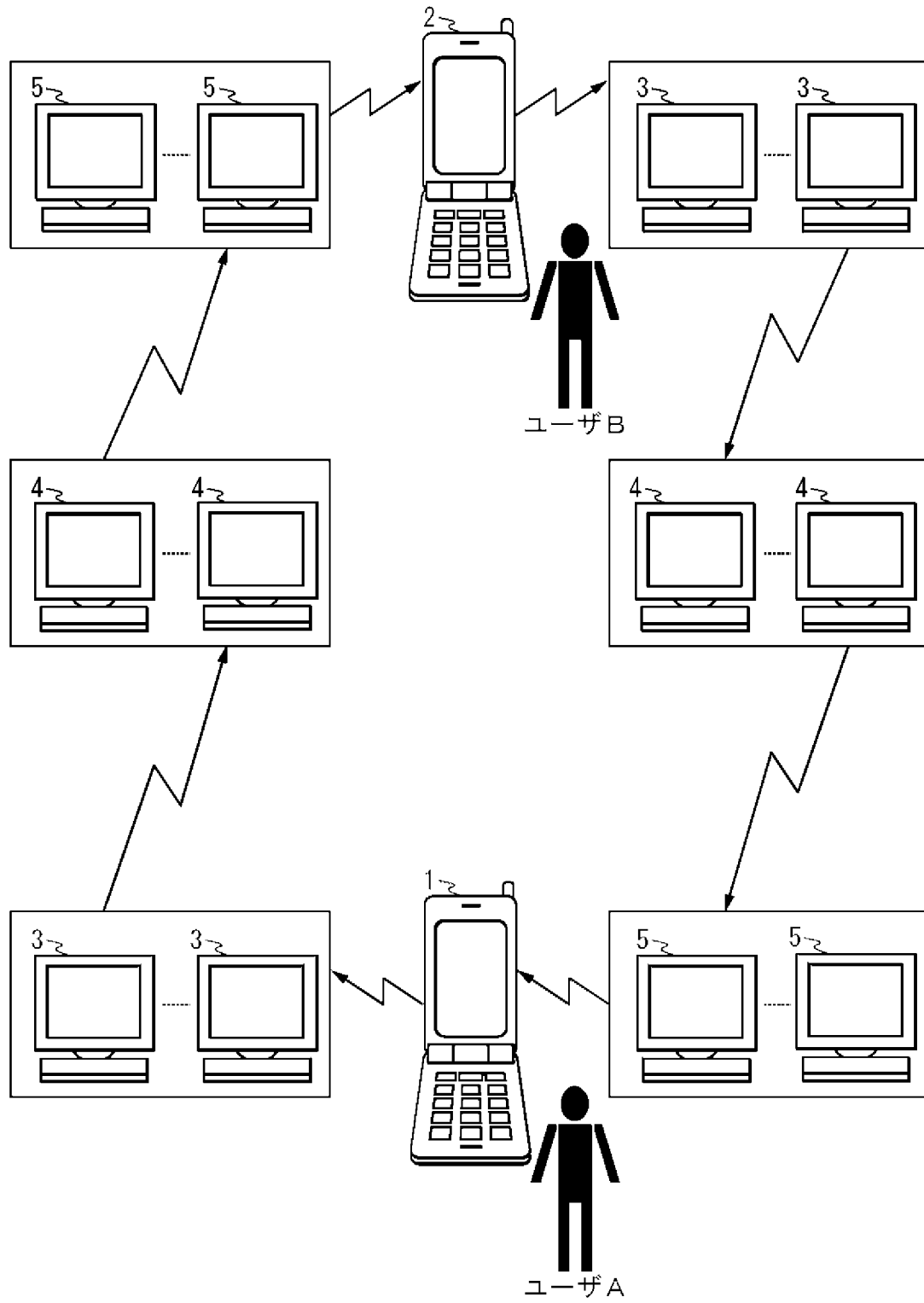
[請求項11] 請求項 1 記載の音声翻訳システムを構成する第一端末装置。

[請求項12] 請求項 2 または請求項 3 記載の音声翻訳システムを構成する音声認識サーバ装置。

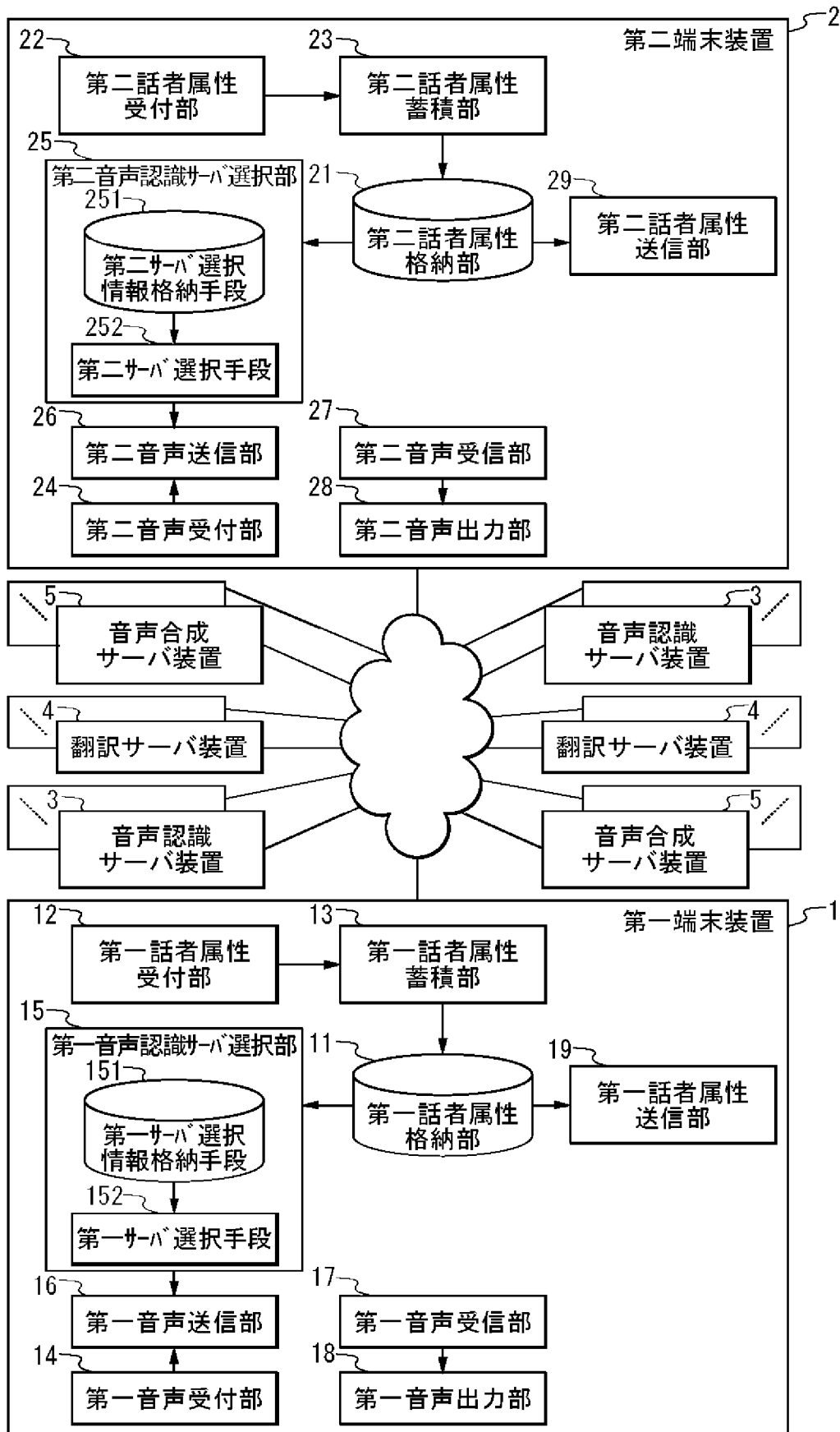
[請求項13] 請求項 4 または請求項 5 記載の音声翻訳システムを構成する翻訳サーバ装置。

[請求項14] 請求項 6 記載の音声翻訳システムを構成する音声合成サーバ装置。

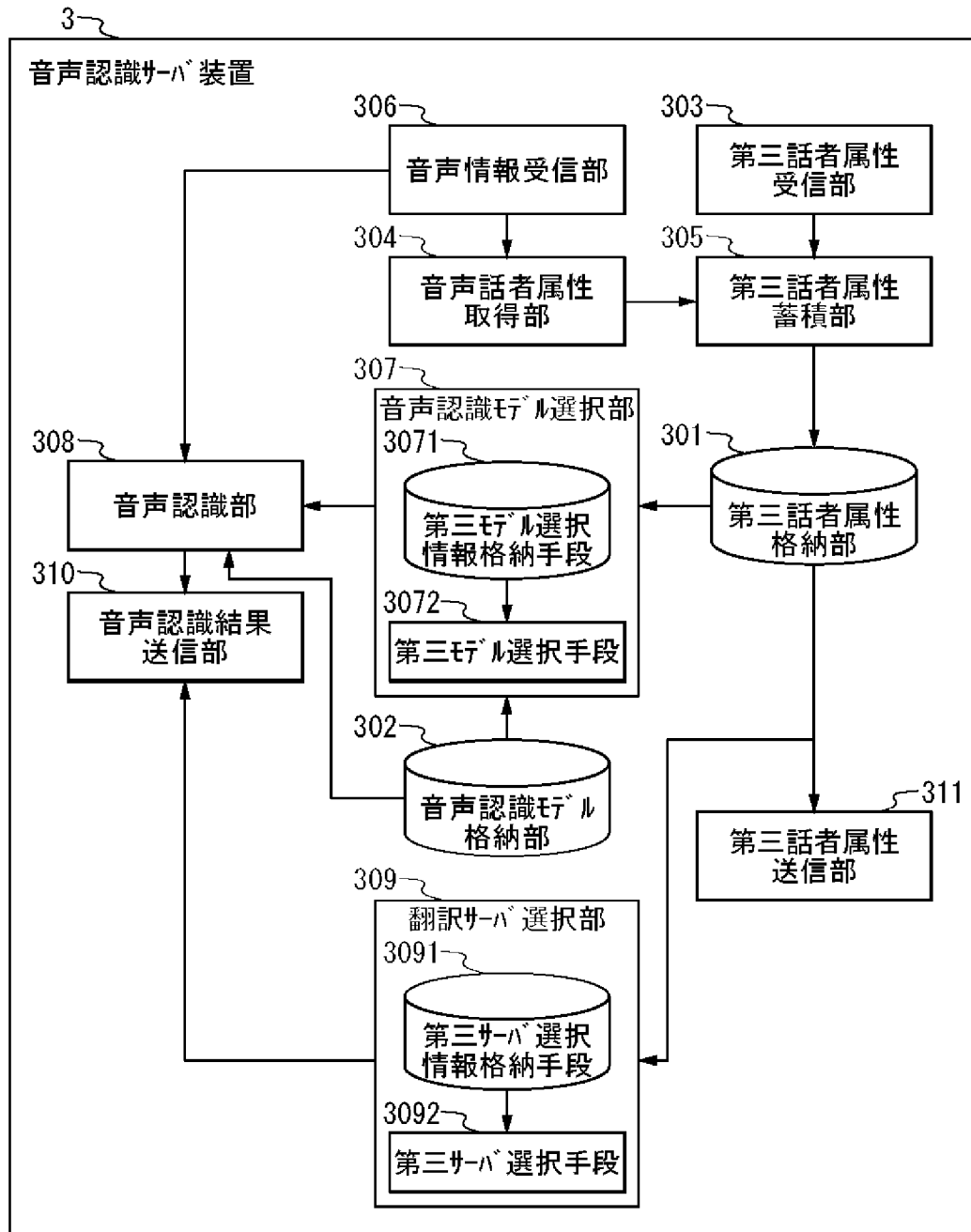
[図1]



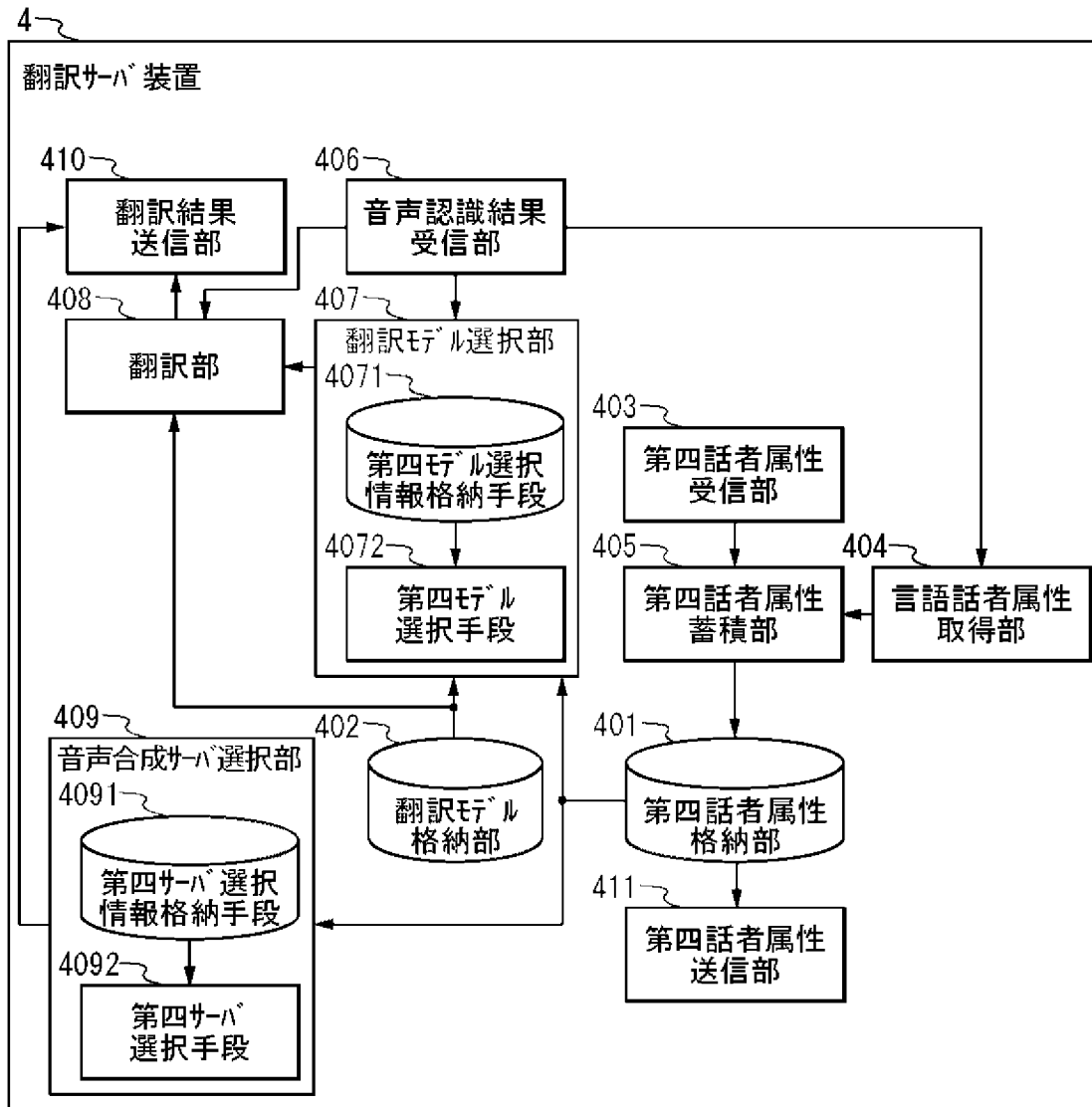
[図2]



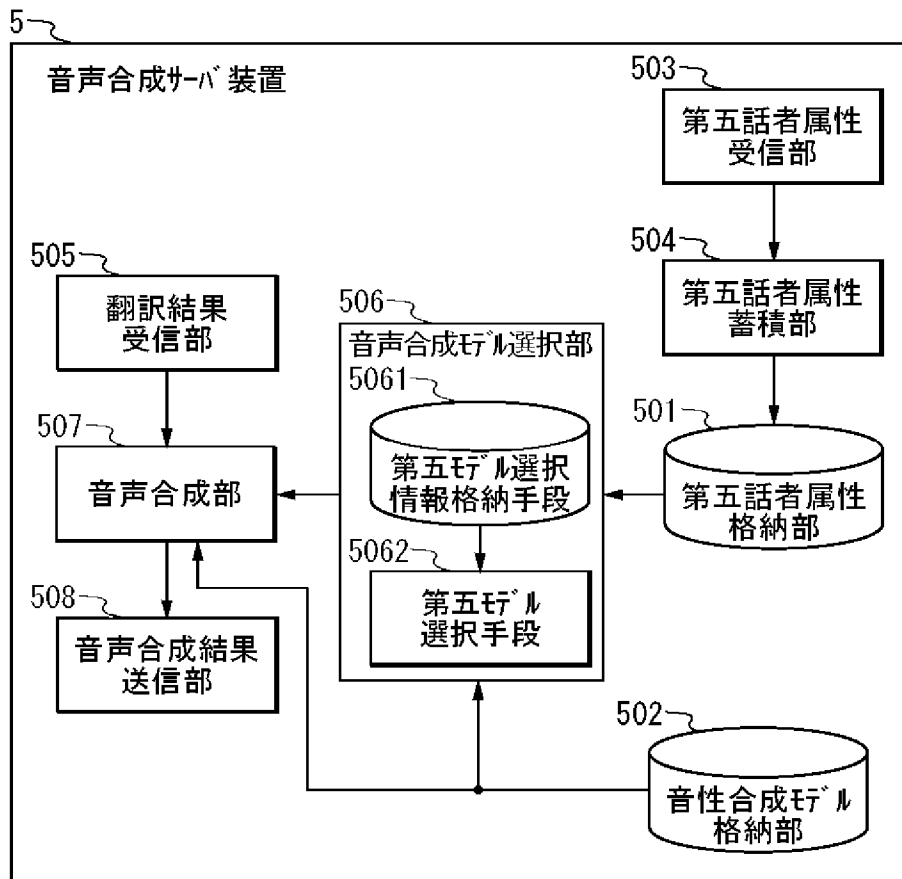
[図3]



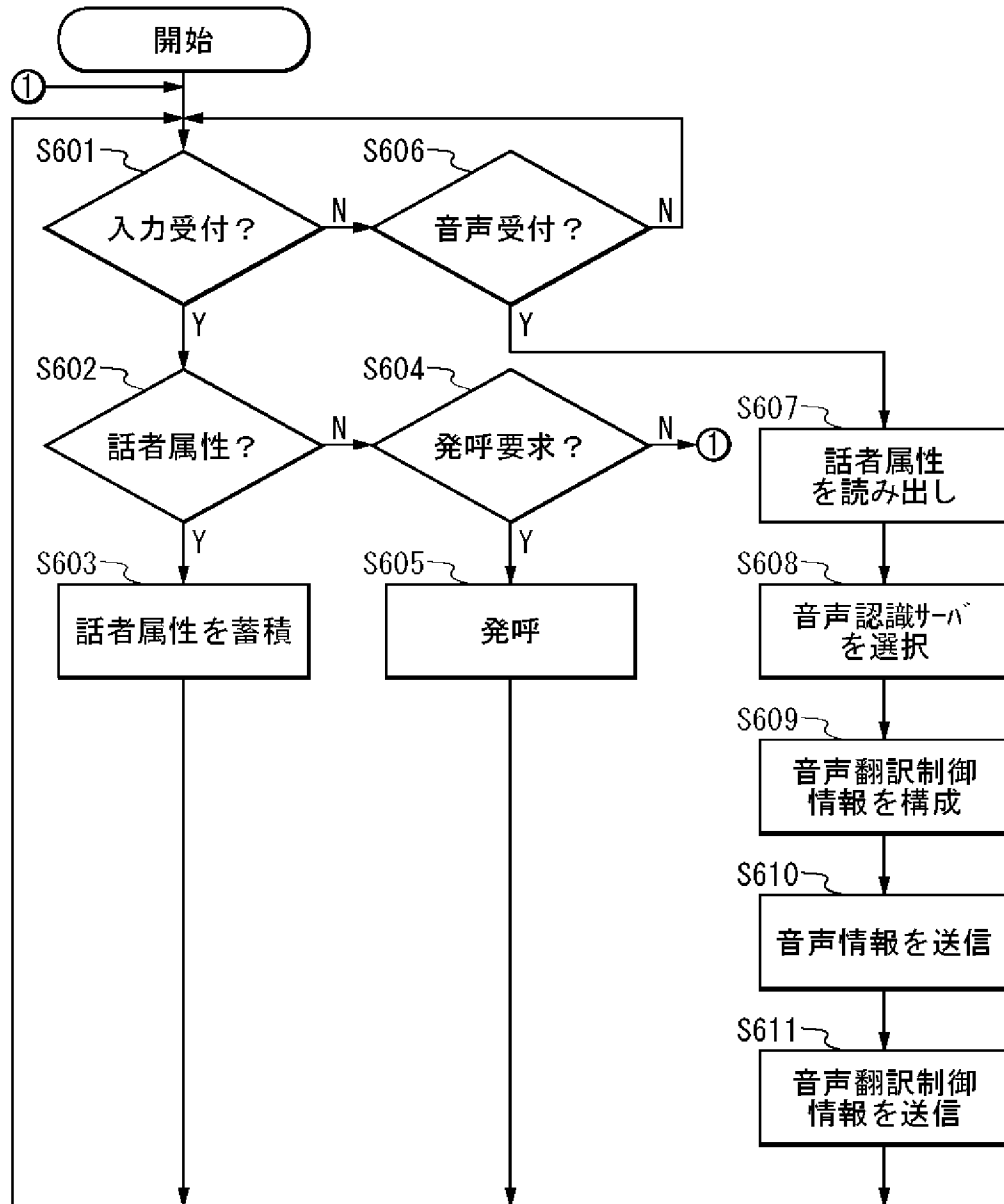
[図4]



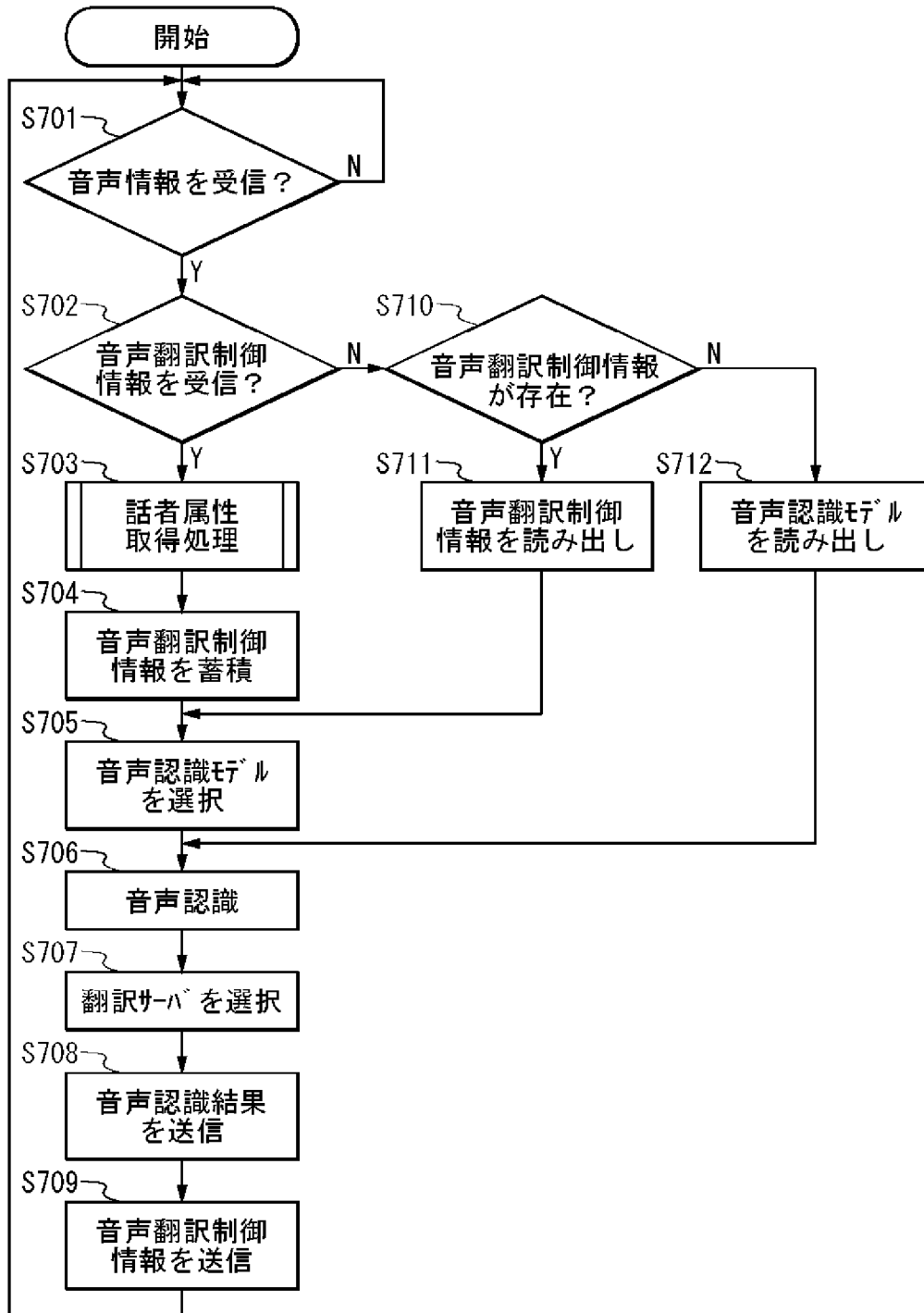
[図5]



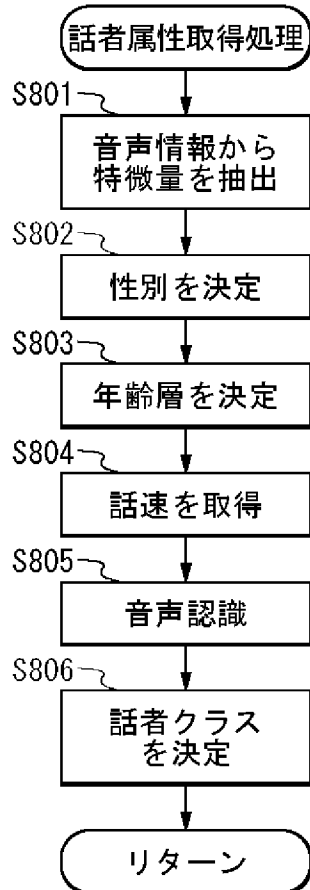
[図6]



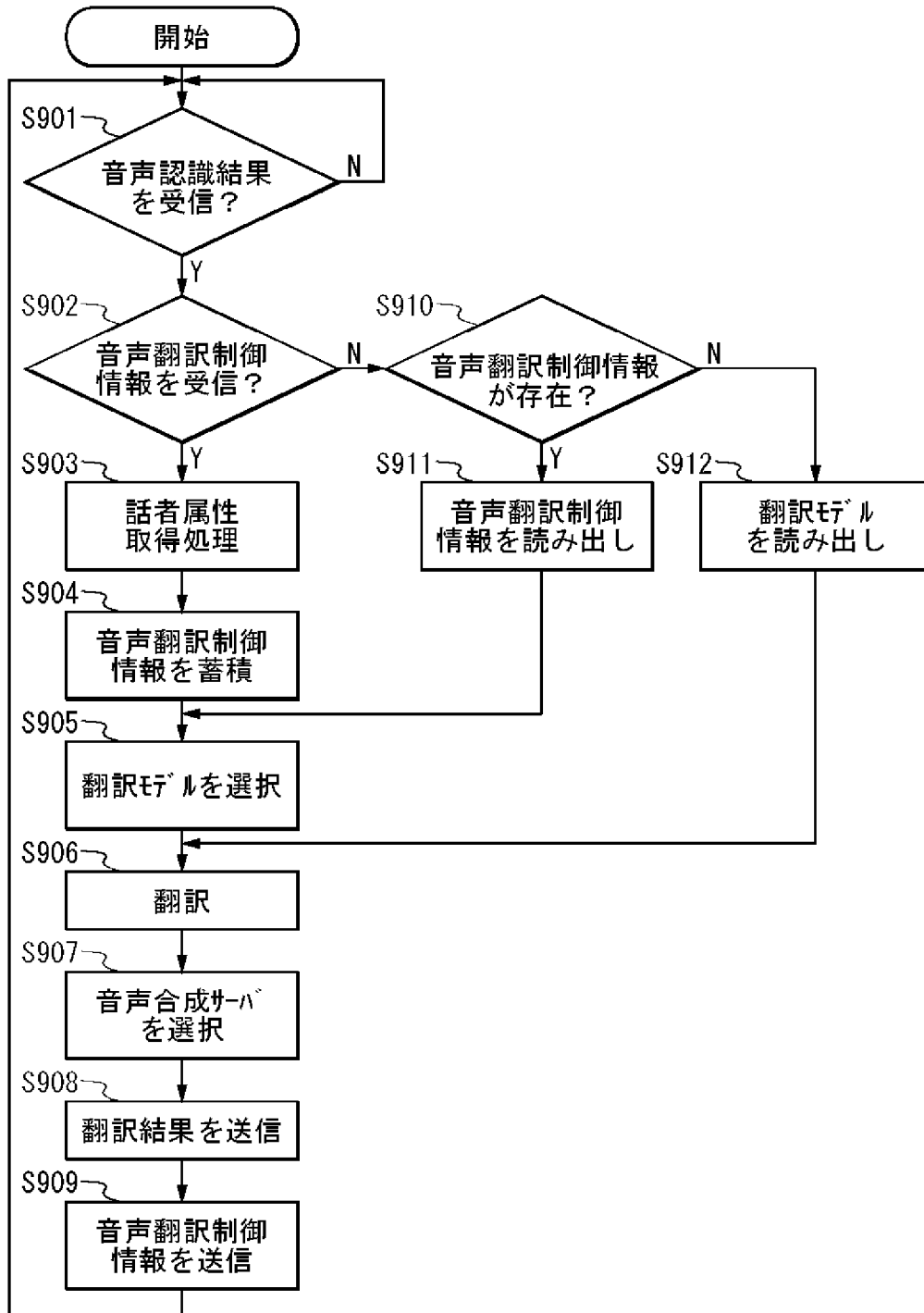
[図7]



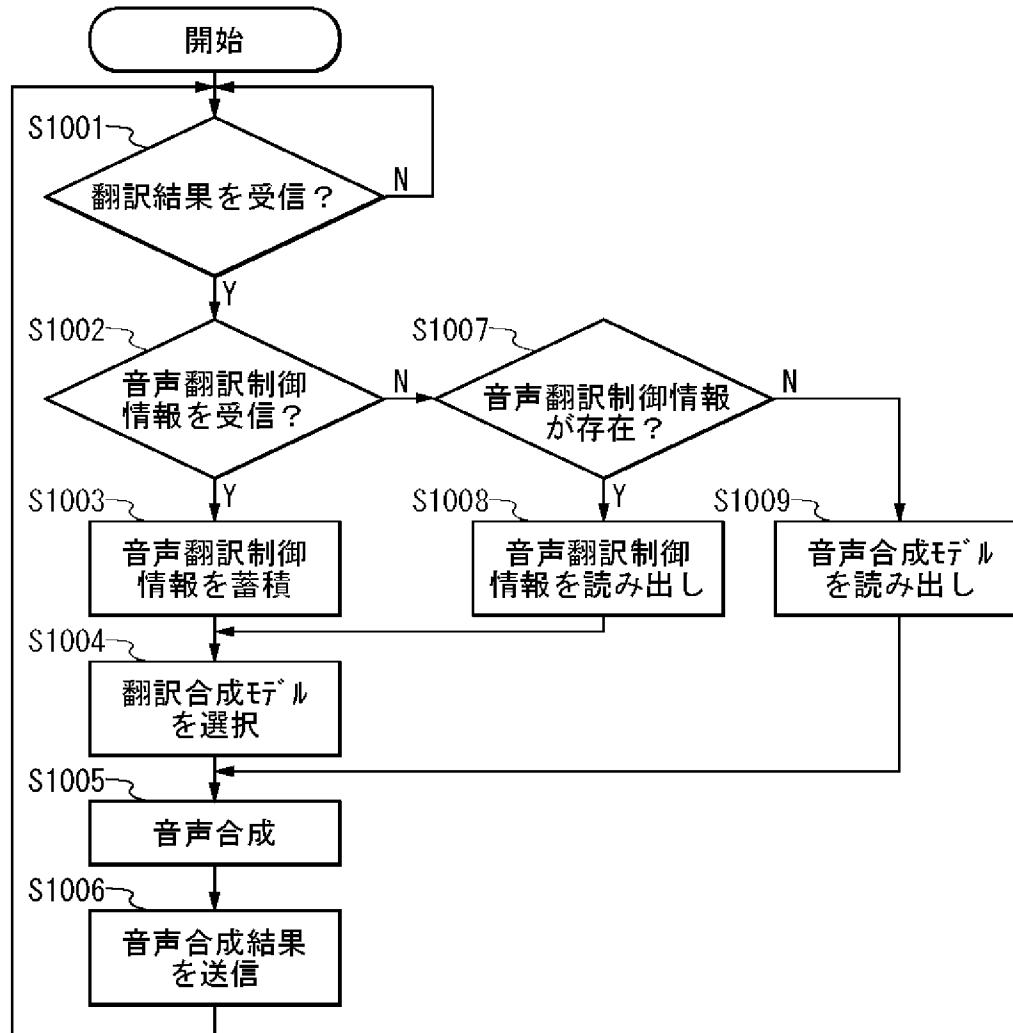
[図8]



[図9]



[図10]



[図11]

性別	女性
年齢	37歳
使用言語	日本語
Native	yes

[図12]

性別	男性
年齢	38歳
使用言語	英語
Native	yes

[図13]

ID	言語	話者属性			音声認識サーバ装置識別子
		性別	年齢	．．．．	
1	日本語	男性	0～9歳	．．．．	254.221.21.39
			10～19歳	．．．．	
2	日本語	男性	20～29歳	．．．．	254.221.29.128
			30～39歳	．．．．	
			40～49歳	．．．．	
3	日本語	男性	50～59歳	．．．．	254.221.38.209
			60～69歳	．．．．	
			70歳以上	．．．．	
4	日本語	女性	0～9歳	．．．．	186.221.57.88
5	日本語	女性	10～19歳	．．．．	186.221.58.129
6	日本語	女性	20～29歳	．．．．	186.221.99.198
7	日本語	女性	30～39歳	．．．．	186.221.1.27
⋮	⋮	⋮	⋮	⋮	⋮
32	英語	男性	0～9歳	．．．．	98.55.30.221
33	英語	男性	10～19歳	．．．．	98.55.77.125
34	英語	男性	20～29歳	．．．．	98.55.208.120
35	英語	男性	30～39歳	．．．．	98.55.173.217
⋮	⋮	⋮	⋮	⋮	⋮

[図14]

ID	言語	話者属性				音声認識エラー識別子
		性別	年齢	話速		
.....						
17	日本語	女性	30～39歳	速い		JR5
18	日本語	女性	30～39歳	中程度		JR6
19	日本語	女性	30～39歳	遅い		JR7
20	日本語	女性	40～49歳	速い		JR8
.....						
39	英語	男性	30～39歳	速い		ER3
40	英語	男性	30～39歳	中程度		ER4
41	英語	男性	30～39歳	遅い		ER5
42	英語	男性	40～49歳	速い		ER6
.....						

ID	原言語	目的言語	話者属性				翻訳サーバ装置識別子
			性別	年齢	第一話者クラス		
1							
25	日本語	英語	女性	30～39歳	A	77.128.50.80	
26	日本語	英語	女性	30～39歳	B	77.128.55.58	
27	日本語	英語	女性	30～39歳	C	77.128.29.107	
3							
38	英語	日本語	男性	30～39歳	A	187.25.30.8	
39	英語	日本語	男性	30～39歳	B	187.25.38.29	
40	英語	日本語	男性	30～39歳	C	187.25.58.227	
4							

[図16]

ID	原言語	話者属性				翻訳モデル識別子
		性別	年齢	第二話者クラス		
.....						
17	日本語	女性	20～29歳	N		JT3
18	日本語	女性	30～39歳	Y		JT4
19	日本語	女性	30～39歳	N		JT5
20	日本語	女性	40～49歳	Y		JT6
.....						
39	英語	男性	20～29歳	N		ET6
40	英語	男性	30～39歳	Y		ET7
41	英語	男性	30～39歳	N		ET8
42	英語	男性	40～49歳	Y		ET9
.....						

[図17]

ID	目的言語	話者属性						音声合成サーバ装置識別子
		性別	年齢	話速	第一話者クラス	第二話者クラス		
1	英語	女性	30～39歳	中程度	A	Y	238.3.55.7	
33	英語	女性	30～39歳	中程度	A	N	238.3.88.101	
34	英語	女性	30～39歳	中程度	B	Y	238.3.89.51	
35	英語	女性	30～39歳	中程度	B	Y	238.3.89.51	
58	日本語	男性	30～39歳	速い	A	Y	239.52.7.51	
59	日本語	男性	30～39歳	速い	A	N	239.52.81.9	
60	日本語	男性	30～39歳	速い	B	Y	239.52.108.9	
1	英語	女性	30～39歳	中程度	A	Y	238.3.55.7	

[図18]

ID	目的言語	話者属性				音声合成時の識別子
		性別	年齢	第二話者クラス		
1	英語	女性	20～29歳	N		JC8
17	英語	女性	30～39歳	Y		JC9
18	英語	女性	30～39歳	N		JC10
19	英語	女性	40～49歳	Y		JC11
20	英語	女性	20～29歳	N		EC1
39	日本語	男性	30～39歳	Y		EC2
40	日本語	男性	30～39歳	N		EC3
41	日本語	男性	40～49歳	Y		EC4
42	日本語	男性	20～29歳	N		EC1
43	日本語	女性	30～39歳	Y		EC2
44	日本語	女性	30～39歳	N		EC3
45	日本語	女性	40～49歳	Y		EC4

[図19]

Figure 19 illustrates a mobile phone interface for a language learning application. The screen displays a form with the following sections:

- 話者属性 (Talker Attributes):**
 - 性別 (Gender): ☐ 男性 (Male) ☒ 女性 (Female)
 - 年齢 (Age): ☐ 0~9才 ☐ 10~19才 ☐ 20~29才 ☒ 30~39才 ☐ 40~49才 ☐ 50~59才 ☐ 60~69才 ☐ 70才以上
- 使用言語 (Languages Used):**
 - ☐ 英語 (English) ☒ 日本語 (Japanese) ☐ 中国語 (Chinese)
 - ☐ ハンゲル語 (Hungarian) ☐ フランス語 (French) -----
- Native:**
 - ☒ Yes ☐ No
- 相手の使用言語 (Partner's Languages):**
 - ☒ 英語 (English) ☐ 日本語 (Japanese) ☐ 中国語 (Chinese)
 - ☐ ハンゲル語 (Hungarian) ☐ フランス語 (French) -----
- 自分の電話番号 (Own Phone Number):** 080-1111-2256
- 相手の電話番号 (Partner's Phone Number):** 092-1445-1122
- 送信 (Send):** (button)

Below the screen are three physical buttons and a circular navigation pad.

[図20]

<原言語> 日本語
 <目的言語> 英語
 <音声認識サーバ> 186.221.1.27
 <相手端末> 092-1445-1122
 <自端末> 080-1111-2256
 <性別> 女性
 <年齢> 30~39
 <第二話者クラス> Y

[図21]

<原言語> 日本語
 <目的言語> 英語
 <音声認識サーバ> 186.221.1.27
 <相手端末> 092-1445-1122
 <自端末> 080-1111-2256
 <性別> 女性
 <年齢> 30~39
 <第二話者クラス> Y
 <話速> 中程度
 <第一話者クラス> A

[図22]

<原言語> 日本語
 <目的言語> 英語
 <音声認識サーバ> 186.221.1.27
 <相手端末> 092-1445-1122
 <自端末> 080-1111-2256
 <性別> 女性
 <年齢> 30~39
 <第二話者クラス> Y
 <話速> 中程度
 <第一話者クラス> A
 <翻訳サーバ> 77.128.50.80

[図23]

＜原言語＞ 日本語
＜目的言語＞ 英語
＜音声認識サーバ＞ 186.221.1.27
＜相手端末＞ 092-1445-1122
＜自端末＞ 080-1111-2256
＜性別＞ 女性
＜年齢＞ 30～39
＜第二話者クラス＞ Y
＜話速＞ 中程度
＜第一話者クラス＞ A
＜翻訳サーバ＞ 77.128.50.80
＜音声合成サーバ＞ 238.3.55.7

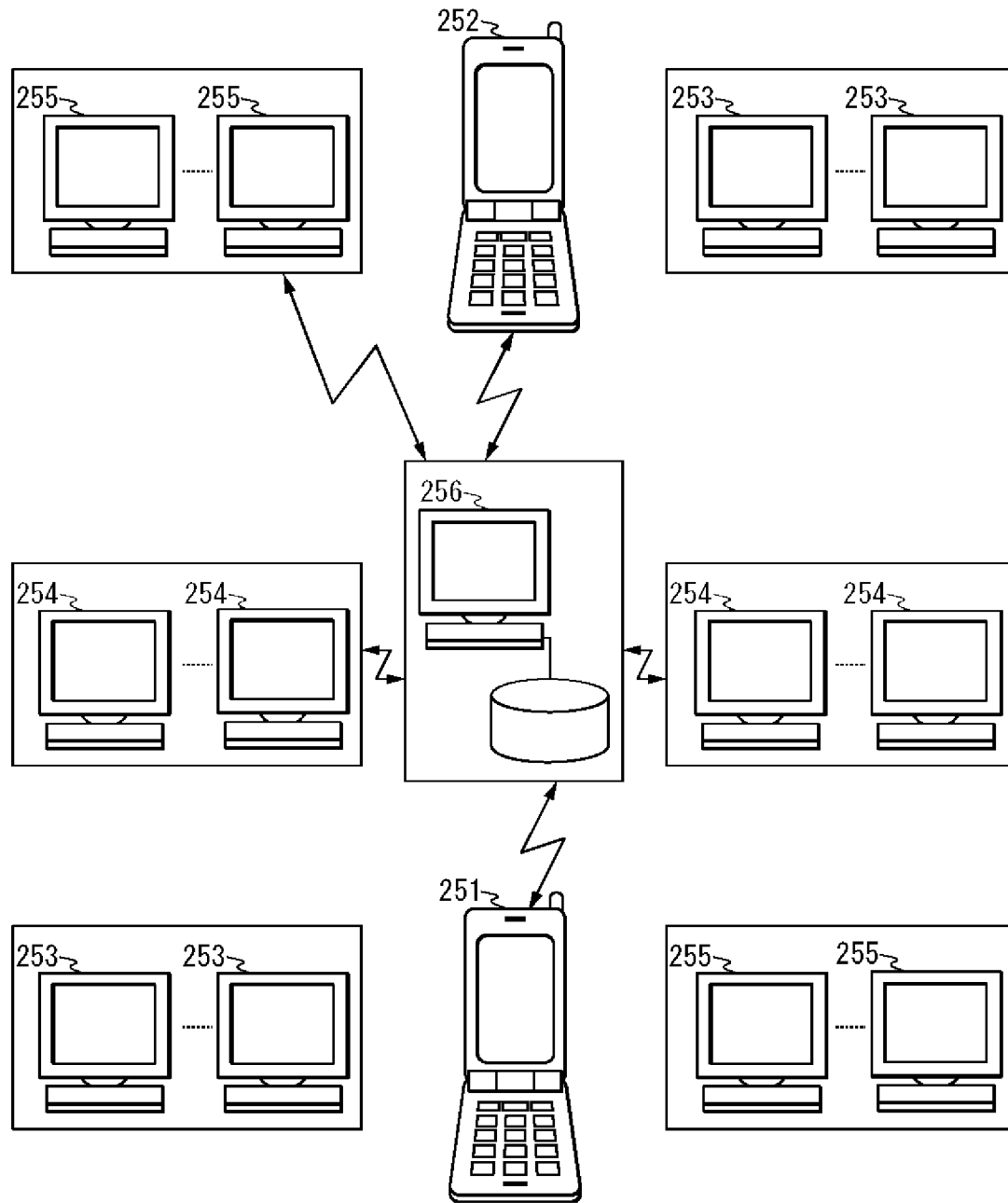
⋮

[図24]

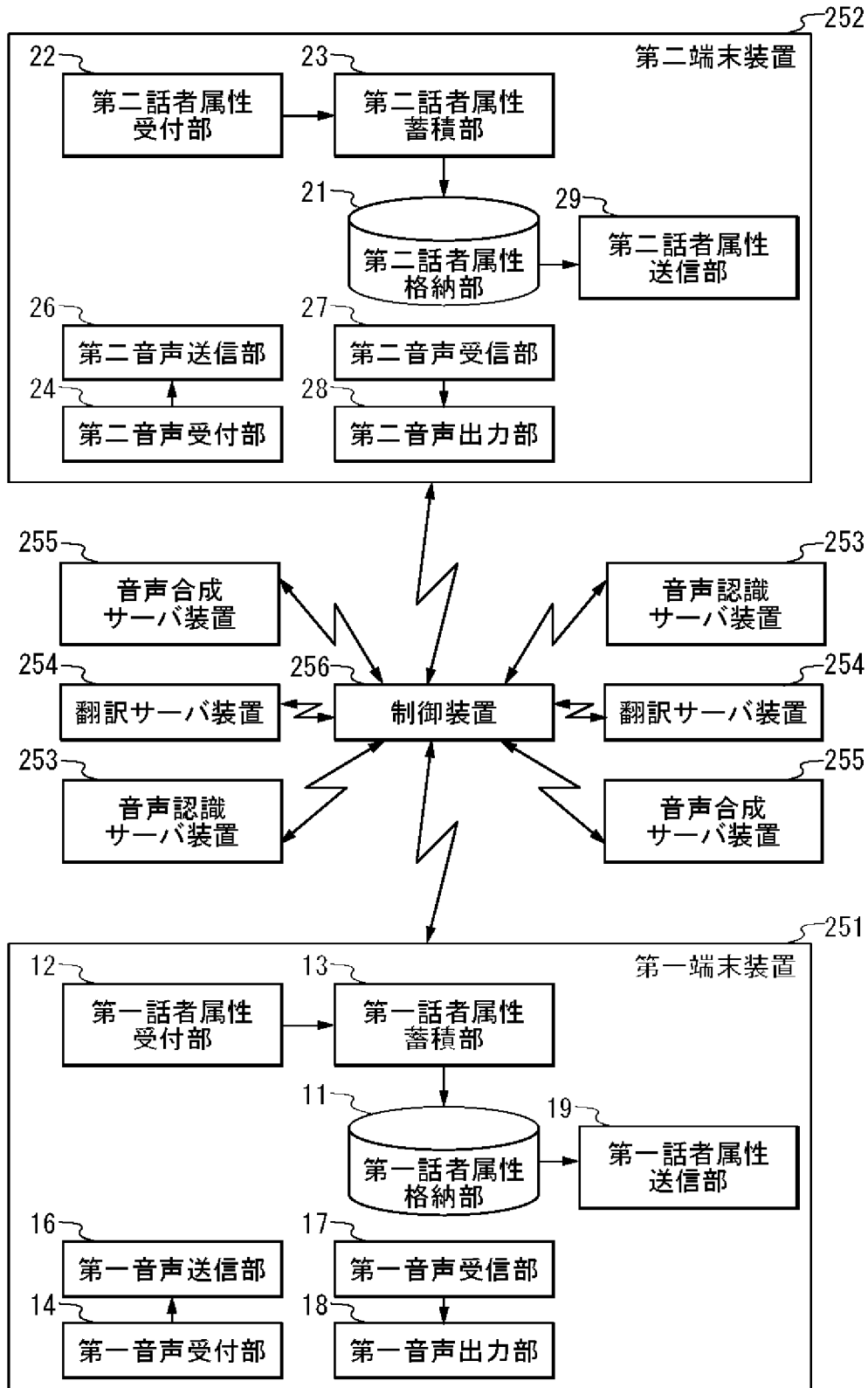
＜原言語＞ 日本語
＜目的言語＞ 英語
＜音声認識サーバ＞ 186.221.1.27
＜翻訳サーバ＞ 225.68.21.129
＜音声合成サーバ＞ 88.125.6.218
＜相手端末＞ 092-1445-1122
＜自端末＞ 080-1111-2256
＜性別＞ 女性
＜年齢＞ 30～39
＜第二話者クラス＞ Y

⋮

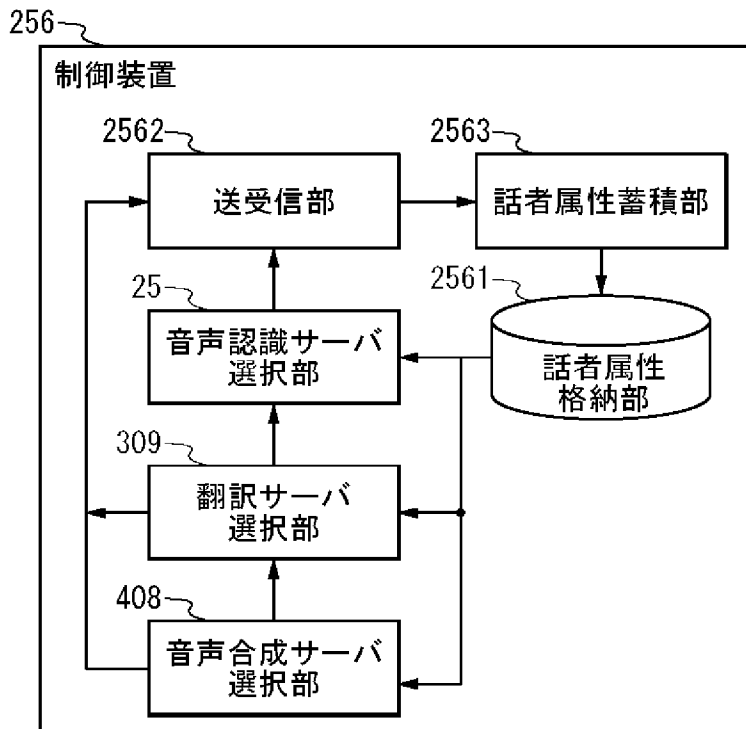
[図25]



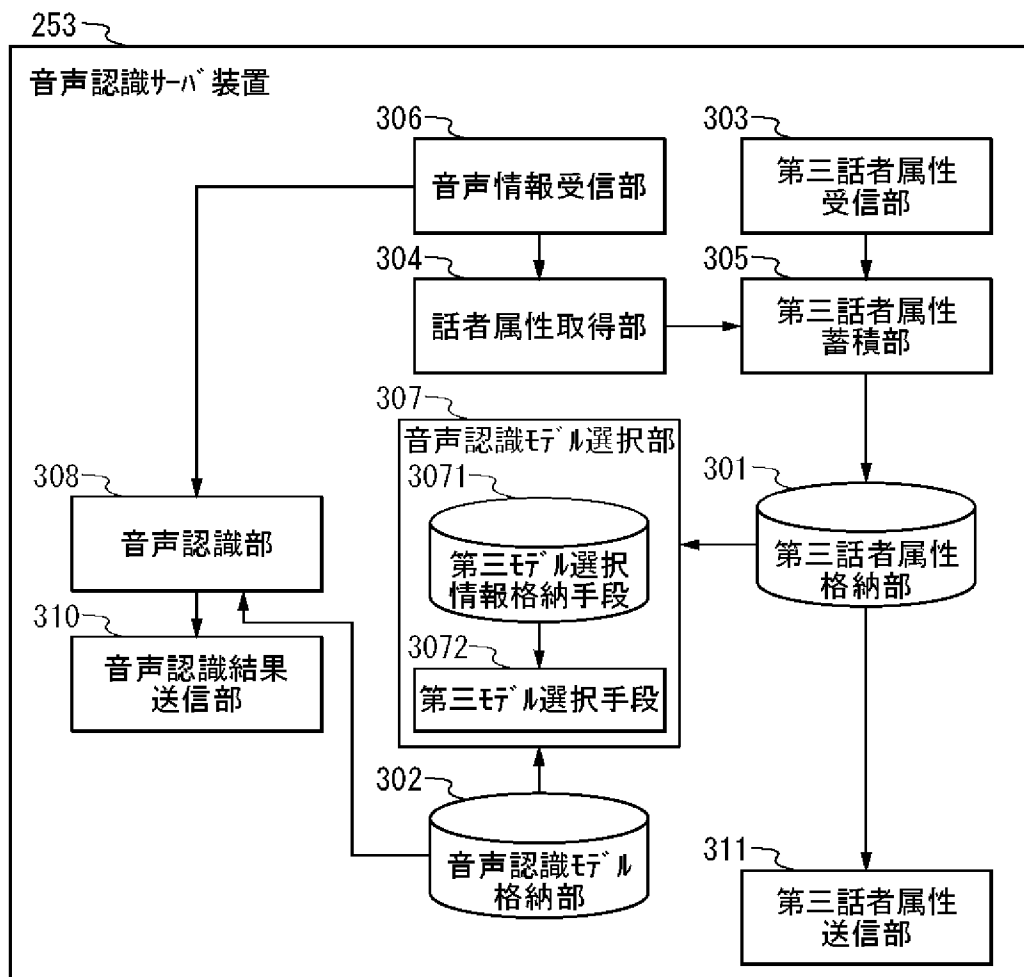
[図26]



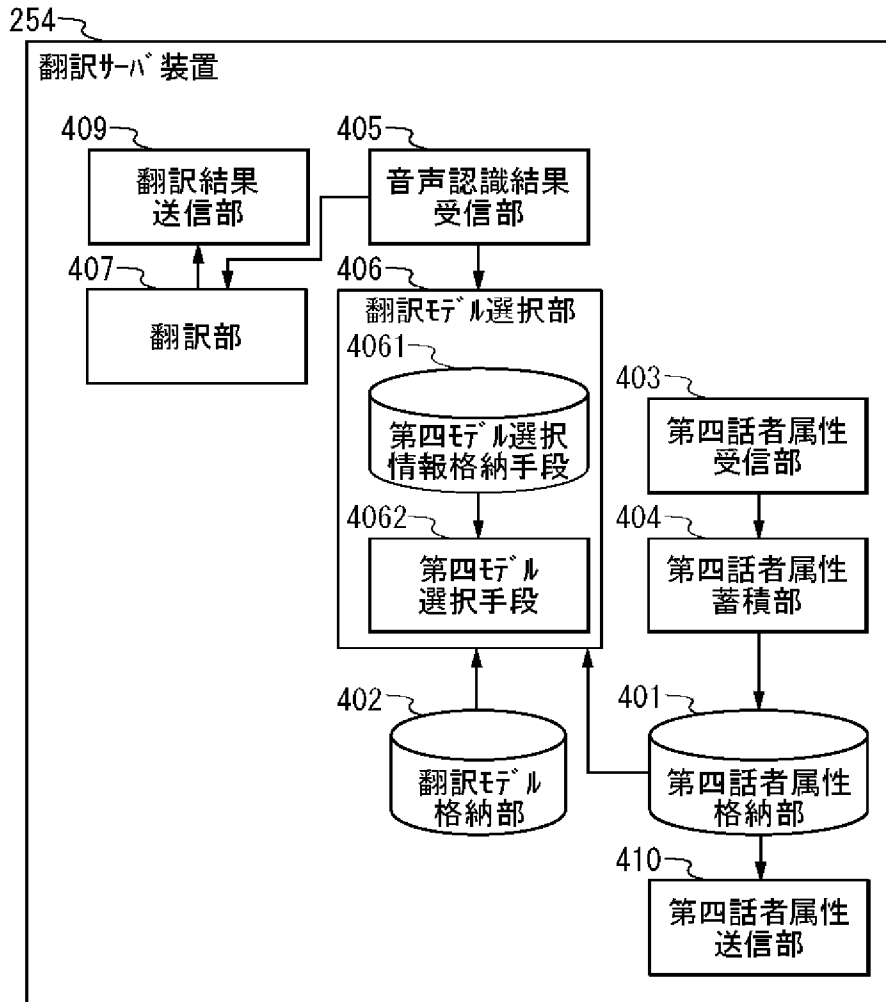
[図27]



[図28]



[図29]



[図30]

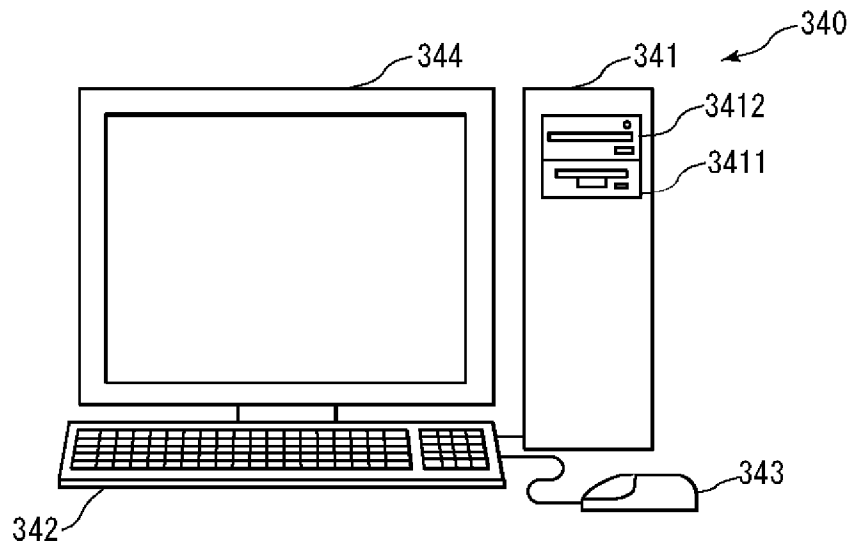
```

<?xml version="1.0" encoding="UTF-8"?>
<STML version="1.0" UtteranceID="8373">
<User ID="Mike"/>
<SR_IN MaxNBest="2" language="en" Task="Dictation" Domain="Travel">
<InputAudioFormat Audio="ADPCM" Endian="Big"/>
<Voice Gender="male" Age="30" Native="no"/>
<OutputTextFormat Form="SurfaceForm"/>

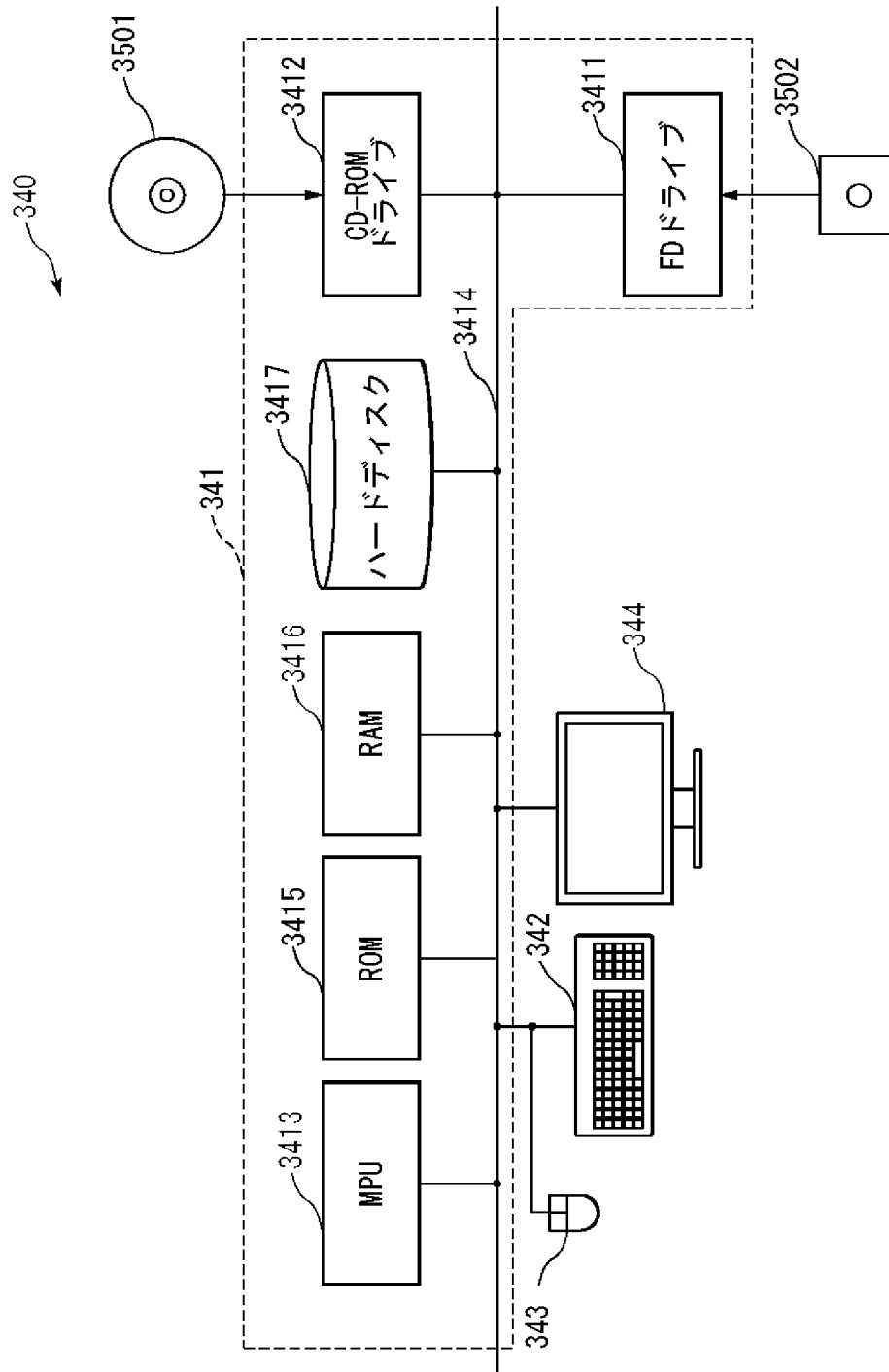
-----
</SR_IN>
</STML>

```

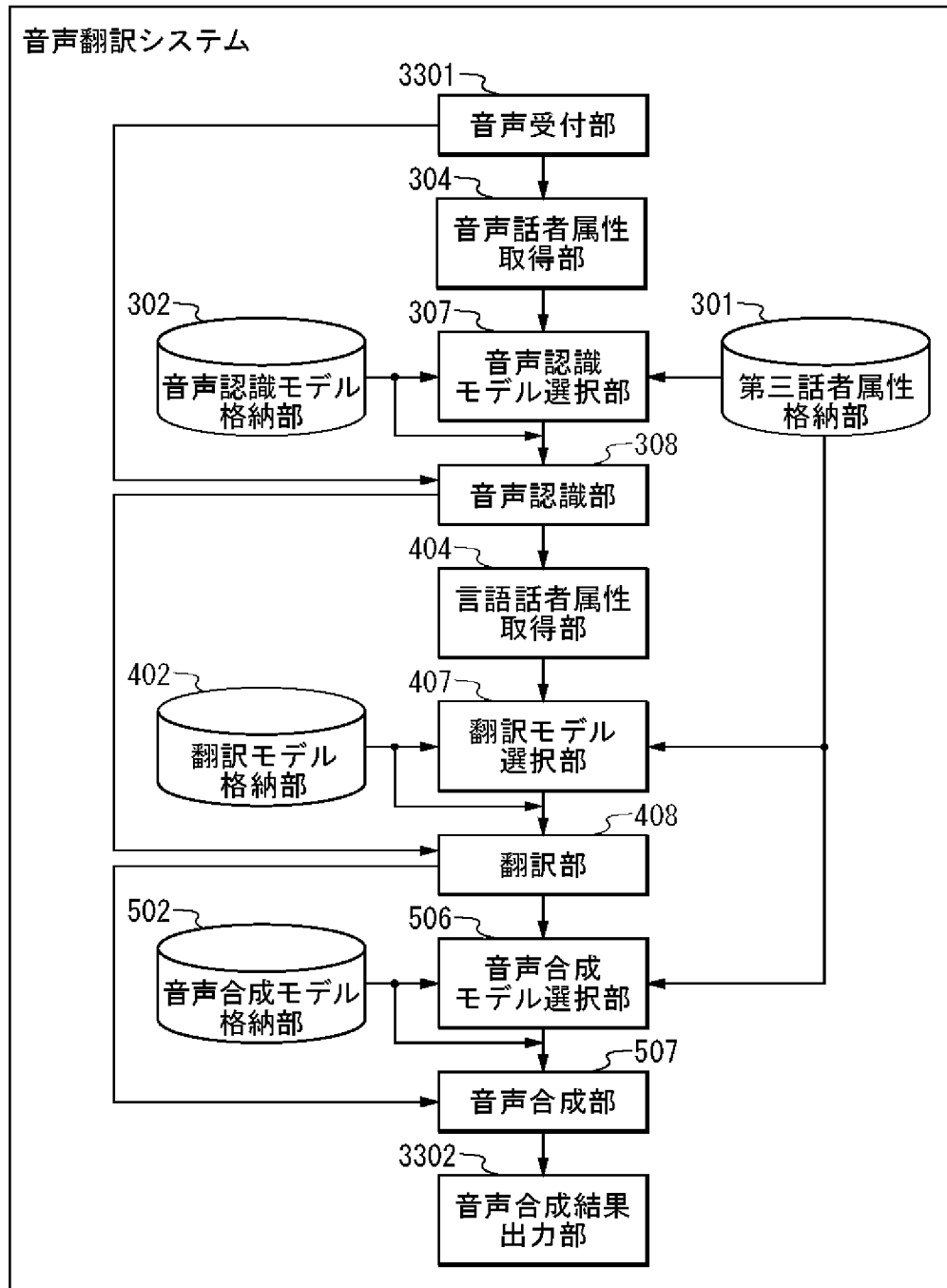
[図31]



[図32]



[図33]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2010/053419

A. CLASSIFICATION OF SUBJECT MATTER

G10L15/10(2006.01)i, G06F3/16(2006.01)i, G06F17/28(2006.01)i, G10L13/00(2006.01)i, G10L15/00(2006.01)i, G10L15/22(2006.01)i, G10L15/28(2006.01)i

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)

G10L15/00-15/28, G06F3/16, G06F17/28, G10L13/00-13/08

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

Jitsuyo Shinan Koho	1922-1996	Jitsuyo Shinan Toroku Koho	1996-2010
Kokai Jitsuyo Shinan Koho	1971-2010	Toroku Jitsuyo Shinan Koho	1994-2010

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y A	JP 2000-148176 A (Sony Corp.), 26 May 2000 (26.05.2000), paragraphs [0029] to [0073]; fig. 1 to 5 (Family: none)	1, 2, 5-14 3, 4
Y A	JP 2005-140988 A (Canon Inc.), 02 June 2005 (02.06.2005), claims 9, 11 (Family: none)	1, 2, 7, 8, 10-12 3-6, 9, 13, 14
Y A	JP 2002-311983 A (ATR Spoken Language Translation Research Laboratories), 25 October 2002 (25.10.2002), claim 1 (Family: none)	5, 6, 9, 13, 14 1-4, 7, 8, 10-12

☒ Further documents are listed in the continuation of Box C.

☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"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)

"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

"T" 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

"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

"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

"&" document member of the same patent family

Date of the actual completion of the international search
30 April, 2010 (30.04.10)

Date of mailing of the international search report
18 May, 2010 (18.05.10)

Name and mailing address of the ISA/
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2010/053419

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2005-31758 A (Canon Inc.), 03 February 2005 (03.02.2005), paragraphs [0036], [0101] & US 2005/0010422 A1	1-14
A	JP 2009-527818 A (Google Inc.), 30 July 2009 (30.07.2009), claim 41 & US 2008/0262828 A1 & EP 1969493 A & WO 2007/098055 A2 & CA 2631815 A & KR 10-2008-0112192 A & AU 2007217900 A	1-14
A	JP 2003-58458 A (Nippon Telegraph And Telephone Corp.), 28 February 2003 (28.02.2003), claim 1 (Family: none)	1-14

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2010/053419

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

The inventions in claims 1 - 14 are common to each other in such a point that those inventions includes "first terminal device", "voice recognition server device", "translation server device", "voice synthesization server device" and "second terminal device". However, since those constituents are well known to a person skilled in the art, this application involves three groups of inventions as follows.

(continued to extra sheet)

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☒ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2010/053419

Continuation of Box No.III of continuation of first sheet (2)

First group: the inventions in claims 1, 2, 7, 8 (referring to claim 2), 10, 11, 12 (referring to claim 2)

Second group: the inventions in claims 3, 4, 8 (referring to claim 3), 9 (referring to claim 4), 12 (referring to claim 3), 13 (referring to claim 4)

Third group: the inventions in claims 5, 6, 9 (referring to claim 5), 13 (referring to claim 5), 14

A. 発明の属する分野の分類（国際特許分類（I P C））

Int.Cl. G10L15/10(2006.01)i, G06F3/16(2006.01)i, G06F17/28(2006.01)i, G10L13/00(2006.01)i,
G10L15/00(2006.01)i, G10L15/22(2006.01)i, G10L15/28(2006.01)i

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（I P C））

Int.Cl. G10L15/00-15/28, G06F3/16, G06F17/28, G10L13/00-13/08

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	1 9 2 2 - 1 9 9 6 年
日本国公開実用新案公報	1 9 7 1 - 2 0 1 0 年
日本国実用新案登録公報	1 9 9 6 - 2 0 1 0 年
日本国登録実用新案公報	1 9 9 4 - 2 0 1 0 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）

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 請求項の番号
Y A	JP 2000-148176 A（ソニー株式会社）2000.05.26, 第0029-0073段落、第1-5図（ファミリーなし）	1, 2, 5-14 3, 4
Y A	JP 2005-140988 A（キャノン株式会社）2005.06.02, 請求項9、11（ファミリーなし）	1, 2, 7, 8, 10-12 3-6, 9, 13, 14

☒ 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

3 0 . 0 4 . 2 0 1 0

国際調査報告の発送日

1 8 . 0 5 . 2 0 1 0

国際調査機関の名称及びあて先

日本国特許庁（I S A / J P）
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）

毛利 太郎

5 Z

3 9 9 0

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 請求項の番号
Y A	JP 2002-311983 A (株式会社エイ・ティ・アール音声言語通信研究所) 2002. 10. 25, 請求項 1 (ファミリーなし)	5, 6, 9, 13, 14 1-4, 7, 8, 10-12
A	JP 2005-31758 A (キャノン株式会社) 2005. 02. 03, 第 0 0 3 6、0 1 0 1 段落 & US 2005/0010422 A1	1-14
A	JP 2009-527818 A (グーグル・インコーポレーテッド) 2009. 07. 30, 請求項 4 1 & US 2008/0262828 A1 & EP 1969493 A & WO 2007/098055 A2 & CA 2631815 A & KR 10-2008-0112192 A & AU 2007217900 A	1-14
A	JP 2003-58458 A (日本電信電話株式会社) 2003. 02. 28, 請求項 1 (ファミリーなし)	1-14

第Ⅱ欄 請求の範囲の一部の調査ができないときの意見（第1ページの2の続き）

法第8条第3項（PCT17条(2)(a)）の規定により、この国際調査報告は次の理由により請求の範囲の一部について作成しなかった。

1. ☐ 請求項 _____ は、この国際調査機関が調査をすることを要しない対象に係るものである。
つまり、
2. ☐ 請求項 _____ は、有意義な国際調査をすることができる程度まで所定の要件を満たしていない国際出願の部分に係るものである。つまり、
3. ☐ 請求項 _____ は、従属請求の範囲であってPCT規則6.4(a)の第2文及び第3文の規定に従って記載されていない。

第Ⅲ欄 発明の単一性が欠如しているときの意見（第1ページの3の続き）

次に述べるようにこの国際出願に二以上の発明があるところこの国際調査機関は認めた。

請求項1－14に係る発明は、「第一端末装置」「音声認識サーバ装置」「翻訳サーバ装置」「音声合成サーバ装置」「第二端末装置」を備える点で共通するものの、これら構成は当業者にとって周知技術に過ぎないから、この出願は次の3群の発明を含むものである。

- 1群：請求項1、2、7、8（請求項2引用）、10、11、12（請求項2引用）に係る発明
- 2群：請求項3、4、8（請求項3引用）、9（請求項4引用）、12（請求項3引用）、13（請求項4引用）に係る発明
- 3群：請求項5、6、9（請求項5引用）、13（請求項5引用）、14に係る発明

1. ☐ 出願人が必要な追加調査手数料をすべて期間内に納付したので、この国際調査報告は、すべての調査可能な請求項について作成した。
2. ☒ 追加調査手数料を要求するまでもなく、すべての調査可能な請求項について調査することができたので、追加調査手数料の納付を求めなかった。
3. ☐ 出願人が必要な追加調査手数料を一部のみしか期間内に納付しなかったため、この国際調査報告は、手数料の納付のあった次の請求項のみについて作成した。
4. ☐ 出願人が必要な追加調査手数料を期間内に納付しなかったため、この国際調査報告は、請求の範囲の最初に記載されている発明に係る次の請求項について作成した。

追加調査手数料の異議の申立てに関する注意

- ☐ 追加調査手数料及び、該当する場合には、異議申立手数料の納付と共に、出願人から異議申立てがあった。
- ☐ 追加調査手数料の納付と共に出願人から異議申立てがあったが、異議申立手数料が納付命令書に示した期間内に支払われなかった。
- ☐ 追加調査手数料の納付はあったが、異議申立てはなかった。