

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2011-22813

(P2011-22813A)

(43) 公開日 平成23年2月3日(2011.2.3)

(51) Int.Cl.	F I	テーマコード (参考)
<b>G06F 17/28 (2006.01)</b>	G06F 17/28 V	5B091
<b>G06F 3/16 (2006.01)</b>	G06F 3/16 340F	5D015
<b>G10L 13/06 (2006.01)</b>	G10L 13/06 250	
<b>G10L 15/00 (2006.01)</b>	G06F 17/28 C	
	G10L 15/00 200C	

審査請求 未請求 請求項の数 12 O L (全 40 頁)

(21) 出願番号 特願2009-167501 (P2009-167501)  
 (22) 出願日 平成21年7月16日 (2009.7.16)

(71) 出願人 301022471  
 独立行政法人情報通信研究機構  
 東京都小金井市貫井北町4-2-1  
 (74) 代理人 100115749  
 弁理士 谷川 英和  
 (72) 発明者 中村 哲  
 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  
 (72) 発明者 隅田 英一郎  
 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  
 (72) 発明者 葦苒 豊  
 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

最終頁に続く

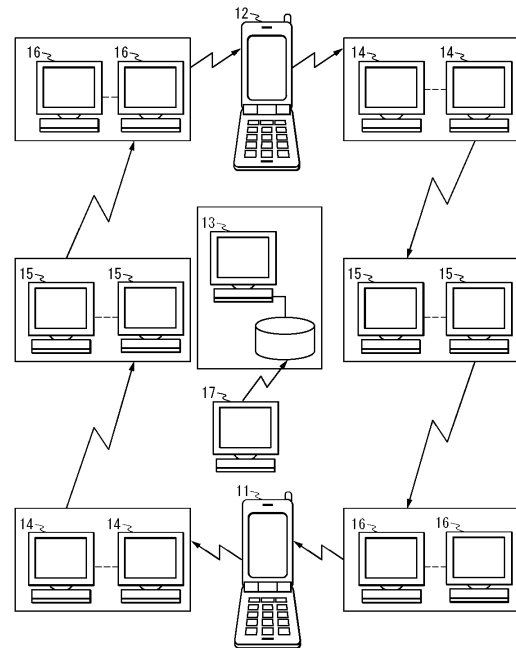
(54) 【発明の名称】 音声翻訳システム、辞書サーバ装置、およびプログラム

(57) 【要約】

【課題】従来、音声翻訳において使用する音声認識辞書、翻訳辞書、および音声合成辞書において不整合があった。

【解決手段】2以上の言語について、音声認識、翻訳、音声合成に利用する全言語用語情報を2以上格納し得る全言語対辞書格納部と、全言語について、用語の音声認識情報を含む音声認識用情報を取得し、1以上の音声認識サーバ装置に送信する音声認識情報送信部と、全言語について、用語の表記を含む翻訳用情報を取得し、1以上の翻訳サーバ装置に送信する翻訳用情報送信部と、全言語について、用語の音声合成情報を含む音声合成用情報を取得し、1以上の音声合成サーバ装置に送信する音声合成用情報送信部とを具備する辞書サーバ装置により、辞書間の不整合を解消できる。

【選択図】 図1



## 【特許請求の範囲】

## 【請求項 1】

辞書サーバ装置と、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、

前記辞書サーバ装置は、

2以上の言語の各言語について、1以上の単語を有する用語の表記と、用語を音声認識するための情報である音声認識情報と、用語を音声合成するための情報である音声合成情報とを有する用語情報を、前記2以上の言語のすべてについて対応付けて有する全言語用語情報を、2以上格納し得る全言語対辞書格納部と、

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、前記用語の音声認識情報を含む音声認識用情報を、前記全言語対辞書格納部から取得し、前記1以上の音声認識サーバ装置に送信する音声認識用情報送信部と、

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、前記用語の表記を含む翻訳用情報を、前記全言語対辞書格納部から取得し、前記1以上の翻訳サーバ装置に送信する翻訳用情報送信部と、

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、前記用語の音声合成情報を含む音声合成用情報を、前記全言語対辞書格納部から取得し、前記1以上の音声合成サーバ装置に送信する音声合成用情報送信部とを具備し、

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、音声認識情報を格納し得る音声認識用情報格納部と、

前記辞書サーバ装置から、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の音声認識用情報を受信する音声認識用情報受信部と、

前記音声認識用情報受信部が受信した音声認識用情報を前記音声認識用情報格納部に蓄積する音声認識用情報蓄積部と、

第一端末装置に入力された音声の情報である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、前記音声認識用情報格納部の音声認識用情報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、翻訳用情報を格納し得る翻訳用情報格納部と、

前記辞書サーバ装置から、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、音翻訳用情報を受信する翻訳用情報受信部と、

前記翻訳用情報受信部が受信した翻訳用情報を前記翻訳用情報格納部に蓄積する翻訳用情報蓄積部と、

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、前記翻訳用情報格納部の翻訳用情報を用いて、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、音声合成用情報を格納し得る音声合成用情報格納部と、

前記辞書サーバ装置から、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、音声合成用情報を受信する音声合成用情報受信部と、

前記音声合成用情報受信部が受信した音声合成用情報を前記音声合成用情報格納部に蓄積する音声合成用情報蓄積部と、

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、前記音声合成用情報格納部の音声合成用情報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、

10

20

30

40

50

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】

前記音声認識サーバ装置の前記音声認識部は、

前記音声情報受信部が受信した音声情報に対する音声認識処理が成功であるか失敗であるかを判断する音声認識判断手段と、

前記音声認識判断手段が、前記音声認識処理が失敗であると判断した場合に、前記辞書サーバ装置に対して、音声認識情報を送信する指示を行う音声認識情報送信指示手段と、

前記音声情報受信部が受信した音声情報を、前記音声認識用情報格納部の音声認識用情報を用いて音声認識し、音声認識結果を取得し、かつ、前記指示の送信に対応して、前記辞書サーバ装置から前記音声認識用情報受信部が受信した音声認識情報を用いて音声認識し、音声認識結果を取得する音声認識手段とを具備し、

前記音声認識用情報受信部は、

前記指示の送信に対応して、前記辞書サーバ装置から前記音声認識情報を受信する請求項 1 記載の音声翻訳システム。

【請求項 3】

前記翻訳サーバ装置の前記翻訳部は、

前記音声認識結果受信部が受信した音声認識結果に対する翻訳処理が成功であるか失敗であるかを判断する翻訳判断手段と、

前記翻訳判断手段が、前記翻訳処理が失敗であると判断した場合に、前記辞書サーバ装置に対して、前記目的言語の用語の表記を送信する指示を行う翻訳情報送信指示手段と、

前記音声認識結果受信部が受信した音声認識結果を、前記翻訳用情報格納部の翻訳用情報を用いて、目的言語に翻訳し、翻訳結果を取得し、かつ、前記指示の送信に対応して、前記辞書サーバ装置から前記翻訳用情報受信部が受信した目的言語の用語の表記を用いて、前記音声認識結果を目的言語に翻訳し、翻訳結果を取得する翻訳手段とを具備し、

前記翻訳用情報受信部は、

前記指示の送信に対応して、前記辞書サーバ装置から前記目的言語の用語の表記を受信する請求項 1 記載の音声翻訳システム。

【請求項 4】

前記音声合成サーバ装置の前記音声合成部は、

前記翻訳結果受信部が受信した翻訳結果に対する音声合成処理が成功であるか失敗であるかを判断する音声合成判断手段と、

前記音声合成判断手段が、前記音声合成処理が失敗であると判断した場合に、前記辞書サーバ装置に対して、音声合成情報を送信する指示を行う音声合成情報送信指示手段と、

前記翻訳結果受信部が受信した翻訳結果を、前記音声合成用情報格納部の音声合成用情報を用いて音声合成し、音声合成結果を取得し、かつ、前記指示の送信に対応して、前記辞書サーバ装置から前記音声合成用情報受信部が受信した音声合成情報を用いて前記翻訳結果を音声合成し、音声合成結果を取得する音声合成手段とを具備し、

前記音声合成用情報受信部は、

前記指示の送信に対応して、前記辞書サーバ装置から前記音声合成情報を受信する請求項 1 記載の音声翻訳システム。

【請求項 5】

前記辞書サーバ装置は、

インターネット上の 1 以上のウェブサーバ装置のウェブページから、前記全言語対辞書格納部に存在しない用語の表記を取得する表記取得部と、

前記表記取得部が取得した用語の表記を、前記全言語対辞書格納部に蓄積する表記蓄積部とをさらに具備する請求項 1 から請求項 4 いずれか記載の音声翻訳システム。

【請求項 6】

前記辞書サーバ装置は、

1 以上の第三端末装置から、用語情報のうちのいずれかの情報を受け付ける情報受付部と

10

20

30

40

50

、  
前記情報受付部が受け付けた情報を、対応する言語の対応する用語の表記に対応付けて、  
前記全言語対辞書格納部に蓄積する情報蓄積部とをさらに具備する請求項 5 記載の音声翻  
訳システム。

【請求項 7】

前記辞書サーバ装置は、

前記全言語用語情報または前記全言語用語情報の一部の情報を出力する出力部をさらに具  
備し、

前記出力部は、

前記全言語用語情報または前記全言語用語情報の一部の情報を出力する場合に、前記 2 以  
上のすべての言語について、予め決められたすべての情報が存在する場合と、予め決めら  
れたすべての情報のうちの一部の情報が存在しない場合とにより、視覚的に異なる態様で  
、前記全言語用語情報または前記全言語用語情報の一部の情報を出力する請求項 5 または  
請求項 6 記載の音声翻訳システム。

10

【請求項 8】

請求項 1 から請求項 7 いずれか記載の音声翻訳システムを構成する辞書サーバ装置。

【請求項 9】

請求項 1 から請求項 7 いずれか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 10】

請求項 1 から請求項 7 いずれか記載の音声翻訳システムを構成する翻訳サーバ装置。

20

【請求項 11】

請求項 1 から請求項 7 いずれか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。

【請求項 12】

記憶媒体に、

2 以上の言語の各言語について、1 以上の単語を有する用語の表記と、用語を音声認識す  
るための情報である音声認識情報と、用語を音声合成するための情報である音声合成情報  
とを有する用語情報を、前記 2 以上の言語のすべてについて対応付けて有する全言語用語  
情報を、2 以上格納しており、

コンピュータを、

前記 2 以上の言語のうちすべての言語または 2 以上の一部の言語について、前記用語の  
音声認識情報を含む音声認識用情報を、前記記憶媒体から取得し、前記 1 以上の音声認識  
サーバ装置に送信する音声認識用情報送信部と、

30

前記 2 以上の言語のうちすべての言語または 2 以上の一部の言語について、前記用語の  
表記を含む翻訳用情報を、前記記憶媒体から取得し、前記 1 以上の翻訳サーバ装置に送信  
する翻訳用情報送信部と、

前記 2 以上の言語のうちすべての言語または 2 以上の一部の言語について、前記用語の  
音声合成情報を含む音声合成用情報を、前記記憶媒体から取得し、前記 1 以上の音声合成  
サーバ装置に送信する音声合成用情報送信部として機能させるためのプログラム。

【発明の詳細な説明】

【技術分野】

40

【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）。

【先行技術文献】

【特許文献】

【0003】

50

【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)

【発明の概要】

【発明が解決しようとする課題】

【0004】

しかしながら、従来、音声翻訳処理を行う際に使用する音声認識辞書、翻訳辞書、および音声合成辞書の間で不整合があり、音声翻訳ができない場合があった。辞書間の不整合とは、ある辞書に存在する用語の情報が他の辞書には存在しない等という現象である。

【課題を解決するための手段】

【0005】

本第一の発明の音声翻訳システムは、辞書サーバ装置と、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、辞書サーバ装置は、2以上の言語の各言語について、1以上の単語を有する用語の表記と、用語を音声認識するための情報である音声認識情報と、用語を音声合成するための情報である音声合成情報を有する用語情報を、2以上の言語のすべてについて対応付けて有する全言語用語情報を、2以上格納し得る全言語対辞書格納部と、2以上の言語のうちすべての言語または2以上の一部の言語について、用語の音声認識情報を含む音声認識用情報を、全言語対辞書格納部から取得し、1以上の音声認識サーバ装置に送信する音声認識用情報送信部と、2以上の言語のうちすべての言語または2以上の一部の言語について、用語の表記を含む翻訳用情報を、全言語対辞書格納部から取得し、1以上の翻訳サーバ装置に送信する翻訳用情報送信部と、2以上の言語のうちすべての言語または2以上の一部の言語について、用語の音声合成情報を含む音声合成用情報を、全言語対辞書格納部から取得し、1以上の音声合成サーバ装置に送信する音声合成用情報送信部とを具備し、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、音声認識用情報を格納し得る音声認識用情報格納部と、辞書サーバ装置から、2以上の言語のうちすべての言語または2以上の一部の言語についての音声認識用情報を受信する音声認識用情報受信部と、音声認識用情報受信部が受信した音声認識用情報を音声認識用情報格納部に蓄積する音声認識用情報蓄積部と、第一端末装置に入力された音声の情報である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、音声認識用情報格納部の音声認識用情報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用情報格納部と、辞書サーバ装置から、2以上の言語のうちすべての言語または2以上の一部の言語について、音翻訳用情報を受信する翻訳用情報受信部と、翻訳用情報受信部が受信した翻訳用情報を翻訳用情報格納部に蓄積する翻訳用情報蓄積部と、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、翻訳用情報格納部の翻訳用情報を用いて、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、音声合成用情報を格納し得る音声合成用情報格納部と、辞書サーバ装置から、2以上の言語のうちすべての言語または2以上の一部の言語について、音声合成用情報を受信する音声合成用情報受信部と、音声合成用情報受信部が受信した音声合成用情報を音声合成用情報格納部に蓄積する音声合成用情報蓄積部と、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、音声合成用情報格納部の音声合成用情報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。

【0006】

かかる構成により、音声翻訳において使用する辞書の間での不整合をなくすることができる。

【0007】

10

20

30

40

50

また、本第二の発明の音声翻訳システムは、第一の発明に対して、音声認識サーバ装置の音声認識部は、音声情報受信部が受信した音声情報に対する音声認識処理が成功であるか失敗であるかを判断する音声認識判断手段と、音声認識判断手段が、音声認識処理が失敗であると判断した場合に、辞書サーバ装置に対して、音声認識情報を送信する指示を行う音声認識情報送信指示手段と、音声情報受信部が受信した音声情報を、音声認識用情報格納部の音声認識用情報を用いて音声認識し、音声認識結果を取得し、かつ、指示の送信に対応して、辞書サーバ装置から音声認識用情報受信部が受信した音声認識情報を用いて音声認識し、音声認識結果を取得する音声認識手段とを具備し、音声認識用情報受信部は、指示の送信に対応して、辞書サーバ装置から音声認識情報を受信する音声翻訳システムである。

10

**【0008】**

かかる構成により、音声翻訳において、音声認識辞書に用語の情報が一時的に存在しない場合でも、音声認識処理の時点で自動的に必要な用語の情報を取得でき、音声認識処理を実行できる。

**【0009】**

また、本第三の発明の音声翻訳システムは、第一の発明に対して、翻訳サーバ装置の翻訳部は、音声認識結果受信部が受信した音声認識結果に対する翻訳処理が成功であるか失敗であるかを判断する翻訳判断手段と、翻訳判断手段が、翻訳処理が失敗であると判断した場合に、辞書サーバ装置に対して、目的言語の用語の表記を送信する指示を行う翻訳情報送信指示手段と、音声認識結果受信部が受信した音声認識結果を、翻訳用情報格納部の翻訳用情報を用いて、目的言語に翻訳し、翻訳結果を取得し、かつ、指示の送信に対応して、辞書サーバ装置から翻訳用情報受信部が受信した目的言語の用語の表記を用いて、音声認識結果を目的言語に翻訳し、翻訳結果を取得する翻訳手段とを具備し、翻訳用情報受信部は、指示の送信に対応して、辞書サーバ装置から目的言語の用語の表記を受信する音声翻訳システムである。

20

**【0010】**

かかる構成により、音声翻訳において、翻訳辞書に用語の情報が一時的に存在しない場合でも、翻訳処理の時点で自動的に必要な用語の情報を取得でき、翻訳処理を実行できる。

**【0011】**

また、本第四の発明の音声翻訳システムは、第一の発明に対して、音声合成サーバ装置の音声合成部は、翻訳結果受信部が受信した翻訳結果に対する音声合成処理が成功であるか失敗であるかを判断する音声合成判断手段と、音声合成判断手段が、音声合成処理が失敗であると判断した場合に、辞書サーバ装置に対して、音声合成情報を送信する指示を行う音声合成情報送信指示手段と、翻訳結果受信部が受信した翻訳結果を、音声合成用情報格納部の音声合成用情報を用いて音声合成し、音声合成結果を取得し、かつ、指示の送信に対応して、辞書サーバ装置から音声合成用情報受信部が受信した音声合成情報を用いて翻訳結果を音声合成し、音声合成結果を取得する音声合成手段とを具備し、音声合成用情報受信部は、指示の送信に対応して、辞書サーバ装置から音声合成情報を受信する音声翻訳システムである。

30

40

**【0012】**

かかる構成により、音声翻訳において、音声合成辞書に用語の情報が一時的に存在しない場合でも、音声合成処理の時点で自動的に必要な用語の情報を取得でき、音声合成処理を実行できる。

**【0013】**

また、本第五の発明の音声翻訳システムは、第一から第四いずれかの発明に対して、辞書サーバ装置は、インターネット上の1以上のウェブサーバ装置のウェブページから、全言語対辞書格納部に存在しない用語の表記を取得する表記取得部と、表記取得部が取得した用語の表記を、全言語対辞書格納部に蓄積する表記蓄積部とをさらに具備する音声翻訳システムである。

50

## 【 0 0 1 4 】

かかる構成により、音声翻訳において使用する辞書の間の不整合をなくすために利用する多国語共通辞書の情報を自動的に補充できる。

## 【 0 0 1 5 】

また、本第六の発明の音声翻訳システムは、第五に対して、辞書サーバ装置は、1以上の第三端末装置から、用語情報のうちのいずれかの情報を受け付ける情報受付部と、情報受付部が受け付けた情報を、対応する言語の対応する用語の表記に対応付けて、全言語対辞書格納部に蓄積する情報蓄積部とをさらに具備する音声翻訳システムである。

## 【 0 0 1 6 】

かかる構成により、音声翻訳において使用する辞書の間の不整合をなくすために利用する多国語共通辞書の情報を補充できる。

10

## 【 0 0 1 7 】

また、本第七の発明の音声翻訳システムは、第五または第六の発明に対して、辞書サーバ装置は、全言語用語情報または全言語用語情報の一部の情報を出力する出力部をさらに具備し、出力部は、全言語用語情報または全言語用語情報の一部の情報を出力する場合に、2以上のすべての言語について、予め決められたすべての情報が存在する場合と、予め決められたすべての情報のうちの一部の情報が存在しない場合とにより、視覚的に異なる態様で、全言語用語情報または全言語用語情報の一部の情報を出力する音声翻訳システムである。

## 【 0 0 1 8 】

20

かかる構成により、音声翻訳において使用する辞書の間の不整合をなくすために利用する多国語共通辞書の情報を補充しやすくできる。

## 【発明の効果】

## 【 0 0 1 9 】

本発明による音声翻訳システムによれば、音声翻訳において使用する辞書の間の不整合をなくすことができる。

## 【図面の簡単な説明】

## 【 0 0 2 0 】

【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

30

【図3】同辞書サーバ装置のブロック図

【図4】同音声認識サーバ装置のブロック図

【図5】同翻訳サーバ装置のブロック図

【図6】同音声合成サーバ装置のブロック図

【図7】同辞書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】同音声合成サーバ装置の動作について説明するフローチャート

【図11】同全言語対辞書を示す図

【図12】同音声認識辞書を示す図

40

【図13】同翻訳辞書を示す図

【図14】同音声合成辞書を示す図

【図15】同音声翻訳制御情報を示す図

【図16】同音声翻訳制御情報を示す図

【図17】同全言語対辞書を示す図

【図18】同出力例を示す図

【図19】同コンピュータシステムの概観図

【図20】同コンピュータシステムのブロック図

【発明を実施するための形態】

## 【 0 0 2 1 】

50

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。

【0022】

(実施の形態1)

本実施の形態において、全言語対辞書を有する辞書サーバ装置を用いて、音声翻訳において使用する辞書間の不整合をなくすることができる音声翻訳システムについて説明する。なお、全言語対辞書とは、音声認識、翻訳、および音声合成に必要な情報を一元管理している辞書である。全言語対辞書は、2以上の全言語用語情報を格納している辞書の情報である。全言語用語情報は、音声翻訳の対象となり得る2以上の言語について、言語ごとに、一つの用語情報を有する情報である。用語情報は、音声認識に必要な情報である音声認識情報と、翻訳に必要な情報である翻訳情報と、音声合成に必要な情報である音声合成情報とを有する情報である。用語情報は、一つの用語に関する情報である。また、用語情報は、言語により、その構造が異なっても良い。また、音声翻訳の対象となり得る2以上の言語は、好ましくは3以上の言語である。

10

【0023】

また、本実施の形態において、音声認識や翻訳や音声合成などの各処理の段階で、各処理を行う装置において、必要な情報が存在しない場合に、リアルタイムに、全言語対辞書を有する辞書サーバ装置から、各装置が処理を行うために必要な情報を取得しにいく処理(リアルタイム補完処理、という)を実行する音声翻訳システムについて説明する。

20

【0024】

また、本実施の形態において、例えば、1以上のウェブサーバから、新しく出現した用語などの情報を、クロールなどして取得し、全言語対辞書を充実させていたり、例えば、不特定多数または特定の多数のユーザから、全言語対辞書に蓄積される情報を受け付けたりする機能を有する辞書サーバ装置について説明する。

【0025】

図1は、本実施の形態における音声翻訳システム1の概念図である。音声翻訳システム1は、1以上の第一端末装置11、1以上の第二端末装置12、辞書サーバ装置13、1以上の音声認識サーバ装置14、1以上の翻訳サーバ装置15、1以上の音声合成サーバ装置16、1以上の第三端末装置17を具備する。

30

【0026】

音声翻訳システム1において、例えば、第一端末装置11のユーザ(ユーザA)である日本人が日本語で「おはよう」と発話した場合、音声認識サーバ装置14は日本語「おはよう」を音声認識する。そして、翻訳サーバ装置15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16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12から、音声「Good morning」が出力される。

【0027】

第一端末装置11および第二端末装置12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11を発話する側の端末、第二端末装置12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。

40

【0028】

また、辞書サーバ装置13は、音声認識サーバ装置14、翻訳サーバ装置15、および音声合成サーバ装置16で利用される情報をすべて保有している。この情報が、上述した全言語対辞書である。

【0029】

さらに、第三端末装置17は、全言語対辞書に情報を追記し、全言語対辞書を充実させるために、情報を入力するための端末である。

50



## 【 0 0 3 0 】

図 2 は、本実施の形態における音声翻訳システム 1 のブロック図である。また、図 3 は、辞書サーバ装置 1 3 のブロック図である。図 4 は、音声認識サーバ装置 1 4 のブロック図である。図 5 は、翻訳サーバ装置 1 5 のブロック図である。図 6 は、音声合成サーバ装置 1 6 のブロック図である。

## 【 0 0 3 1 】

第一端末装置 1 1 は、第一音声受付部 1 1 1、第一音声送信部 1 1 2、第一音声受信部 1 1 3、第一音声出力部 1 1 4 を具備する。

## 【 0 0 3 2 】

第二端末装置 1 2 は、第二音声受付部 1 2 1、第二音声送信部 1 2 2、第二音声受信部 1 2 3、第二音声出力部 1 2 4 を具備する。

10

## 【 0 0 3 3 】

図 3 に示すように、辞書サーバ装置 1 3 は、全言語対辞書格納部 1 3 1、音声認識用情報送信部 1 3 2、翻訳用情報送信部 1 3 3、音声合成用情報送信部 1 3 4、表記取得部 1 3 5、表記蓄積部 1 3 6、情報受付部 1 3 7、情報蓄積部 1 3 8、出力部 1 3 9 を具備する。

## 【 0 0 3 4 】

図 4 に示すように、音声認識サーバ装置 1 4 は、音声認識用情報格納部 1 4 1、音声認識用情報受信部 1 4 2、音声認識用情報蓄積部 1 4 3、音声情報受信部 1 4 4、音声認識部 1 4 5、音声認識結果送信部 1 4 6 を具備する。音声認識部 1 4 5 は、音声認識判断手段 1 4 5 1、音声認識情報送信指示手段 1 4 5 2、音声認識手段 1 4 5 3 を具備する。

20

## 【 0 0 3 5 】

図 5 に示すように、翻訳サーバ装置 1 5 は、翻訳用情報格納部 1 5 1、翻訳用情報受信部 1 5 2、翻訳用情報蓄積部 1 5 3、音声認識結果受信部 1 5 4、翻訳部 1 5 5、翻訳結果送信部 1 5 6 を具備する。翻訳部 1 5 5 は、翻訳判断手段 1 5 5 1、翻訳情報送信指示手段 1 5 5 2、翻訳手段 1 5 5 3 を具備する。

## 【 0 0 3 6 】

図 6 に示すように、音声合成サーバ装置 1 6 は、音声合成用情報格納部 1 6 1、音声合成用情報受信部 1 6 2、音声合成用情報蓄積部 1 6 3、翻訳結果受信部 1 6 4、音声合成部 1 6 5、音声合成結果送信部 1 6 6 を具備する。音声合成部 1 6 5 は、音声合成判断手段 1 6 5 1、音声合成情報送信指示手段 1 6 5 2、音声合成手段 1 6 5 3 を具備する。

30

## 【 0 0 3 7 】

第三端末装置 1 7 は、入力受付部 1 7 1、情報受信部 1 7 2、情報出力部 1 7 3、入力情報送信部 1 7 4 を具備する。

## 【 0 0 3 8 】

第一音声受付部 1 1 1 は、第一端末装置 1 1 のユーザ（ユーザ A という）から音声を受け付ける。第一音声受付部 1 1 1 は、例えば、マイクとそのデバイスドライバなどにより実現され得る。

## 【 0 0 3 9 】

第一音声送信部 1 1 2 は、第一音声受付部 1 1 1 が受け付けた音声を送信する。音声の送信先は、1 以上の音声認識サーバ装置 1 4 のうちのいずれかである。なお、第一音声送信部 1 1 2 は、2 以上の音声認識サーバ装置 1 4 に音声を送信しても良い。また、ここで、音声とは音声の情報であり、送信する音声はデジタル化されていることは好適である。また、第一音声送信部 1 1 2 は、音声とともに、音声翻訳制御情報を送信しても良い。音声翻訳制御情報は、音声認識サーバ装置 1 4、翻訳サーバ装置 1 5、および音声合成サーバ装置 1 6 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 1 や第二端末装置 1 2 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、第一端末装置 1 1 や第二端末装置

40

50

12は、例えば、第二端末装置12や第一端末装置11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、第二端末装置12や第一端末装置11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、第一端末装置11や第二端末装置12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14を特定する情報(IPアドレスなど)や、翻訳サーバ装置15を特定する情報(IPアドレスなど)や、音声合成サーバ装置16を特定する情報(IPアドレスなど)を取得する。つまり、第一端末装置11や第二端末装置12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、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。さらに、第一音声送信部112は、直接的に音声を1以上の音声認識サーバ装置14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14に送信しても良い。第一音声送信部112は、通常、無線または有線の通信手段で実現されるが、放送手段で実現されても良い。

10

**【0040】**

第一音声受信部113は、音声(通常、デジタル化されている音声情報)を受信する。この音声は、第二端末装置12のユーザ(ユーザBという)が発声した音声の内容を、第一端末装置11のユーザAが理解できる言語に翻訳された音声である。第一音声受信部113は、音声合成サーバ装置16から、直接的または間接的に音声を受信する。第一音声受信部113は、通常、無線または有線の通信手段で実現されるが、放送を受信する手段で実現されても良い。

20

**【0041】**

第一音声出力部114は、第一音声受信部113が受信した音声を出力する。第一音声出力部114は、スピーカ(出力デバイス)を含むと考えても含まないと考えても良い。第一音声出力部114は、出力デバイスのドライバソフトまたは、出力デバイスのドライバソフトと出力デバイス等で実現され得る。

**【0042】**

第二音声受付部121は、第二端末装置12のユーザBから音声を受け付ける。第二音声受付部121は、例えば、マイクとそのデバイスドライバなどにより実現され得る。

30

**【0043】**

第二音声送信部122は、第二音声受付部121が受け付けた音声を送信する。音声の送信先は、1以上の音声認識サーバ装置14のうちのいずれかである。なお、第二音声送信部122は、2以上の音声認識サーバ装置14に音声を送信しても良い。また、ここで、音声とは音声の情報であり、送信する音声はデジタル化されていることは好適である。さらに、第二音声送信部122は、直接的に音声を1以上の音声認識サーバ装置14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14に送信しても良い。第二音声送信部122は、通常、無線または有線の通信手段で実現されるが、放送手段で実現されても良い。

40

**【0044】**

第二音声受信部123は、音声(通常、デジタル化されている音声情報)を受信する。この音声は、第一端末装置11のユーザAが発声した音声の内容を、第二端末装置12のユーザBが理解できる言語(目的言語)に翻訳された音声である。第二音声受信部123は、音声合成サーバ装置16から、直接的または間接的に音声を受信する。第二音声受信部123は、通常、無線または有線の通信手段で実現されるが、放送を受信する手段で実現されても良い。

**【0045】**

第二音声出力部124は、第二音声受信部123が受信した音声を出力する。第二音声

50

出力部 1 2 4 は、スピーカ（出力デバイス）を含むと考えても含まないと考えても良い。第一音声出力部 1 1 4 は、出力デバイスのドライバソフトまたは、出力デバイスのドライバソフトと出力デバイス等で実現され得る。

【 0 0 4 6 】

全言語対辞書格納部 1 3 1 は、全言語対辞書を格納している。全言語対辞書については、上述した。なお、全言語対辞書が有する音声認識情報は、例えば、隠れマルコフモデル（HMM）などの音響モデルである。また、全言語対辞書が有する翻訳情報は、例えば、用語の表記である。用語の表記とは、言語の日本語が「大阪」である場合、英語は「Osaka」、中国語は「大阪」である。なお、用語とは、一つの単語、1以上の単語、1以上の文節、文なども含む広い概念である、として捉える。音声合成情報とは、例えば、読みの情報（適宜「読み」という）とアクセントの情報（適宜「アクセント」という）などである。用語情報は、用語の読みを、通常、有する。用語情報は、例えば、表記、読み、アクセント、トーンなどの情報を有する。用語情報は、通常、言語により構造は異なる。例えば、言語がドイツ語の場合、用語情報は、用語に対して男性語であるか女性語であるかを示すフラグ情報を有する。なお、かかるフラグ情報は、日本語や英語などの言語の用語情報には存在しない。

10

【 0 0 4 7 】

また、全言語対辞書は、通常、言語ごとの構造情報を有する。例えば、全言語対辞書は、「日本語<HMM><表記><読み><アクセント>」「英語<HMM><表記><読み><アクセント><トーン>」「ドイツ語<表記><フラグ情報><読み><アクセント>」などである。なお、構造情報は、言語共通で有する構造の情報は一つで管理し、言語特有の情報を示す構造情報のみ、言語ごとに管理しても良い。かかる場合、構造情報は、言語共通の構造情報「<HMM><表記><読み><アクセント>」、「ドイツ語<フラグ情報>」等の情報となる。

20

【 0 0 4 8 】

全言語対辞書格納部 1 3 1 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。全言語対辞書格納部 1 3 1 に各種の情報が記憶される過程は問わない。例えば、記録媒体を介して各種の情報が全言語対辞書格納部 1 3 1 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各種の情報が全言語対辞書格納部 1 3 1 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各種の情報が全言語対辞書格納部 1 3 1 で記憶されるようになってよく、なお、後述する表記取得部 1 3 5、表記蓄積部 1 3 6、情報受付部 1 3 7、情報蓄積部 1 3 8 などにより、全言語対辞書格納部 1 3 1 に情報が蓄積されることは好適である。

30

【 0 0 4 9 】

音声認識用情報送信部 1 3 2 は、全言語対辞書格納部 1 3 1 から音声認識用情報を読み出し、当該音声認識用情報を 1 以上の音声認識サーバ装置 1 4 に送信する。音声認識用情報とは、用語の音声認識情報を含む情報であり、音声認識サーバ装置 1 4 で音声認識のために利用される情報である。音声認識用情報は、2以上の言語のうちすべての言語または一部の言語についての音声認識情報を含む情報である。音声認識用情報は、音声認識情報と同一の情報でも良いし、音声認識情報に他の情報を加えた情報でも良い。また、音声認識用情報送信部 1 3 2 は、音声認識用情報の中で一部に欠けている情報がある用語の音声認識情報は送信しないことは好適である。音声認識用情報送信部 1 3 2 は、通常、無線または有線の通信手段で実現されるが、放送手段で実現されても良い。

40

【 0 0 5 0 】

翻訳用情報送信部 1 3 3 は、全言語対辞書格納部 1 3 1 から翻訳用情報を読み出し、当該翻訳用情報を 1 以上の翻訳サーバ装置 1 5 に送信する。翻訳用情報とは、用語の表記を含む情報であり、翻訳サーバ装置 1 5 で翻訳のために利用される情報である。翻訳用情報は、2以上の言語のうちすべての言語または一部の言語についての表記を含む情報である。翻訳用情報は、表記のみでも良いし、表記に他の情報を加えた情報でも良い。また、翻訳用情報送信部 1 3 3 は、翻訳用情報の中で一部に欠けている情報がある用語の翻訳情

50

報は送信しないことは好適である。つまり、日本語の用語の表記のみしか存在しない用語情報があつた場合、当該用語情報が有する情報は、全く送信しないことは好適である。翻訳情報送信部 133 は、通常、無線または有線の通信手段で実現されるが、放送手段で実現されても良い。

#### 【0051】

音声合成用情報送信部 134 は、全言語対辞書格納部 131 から音声合成用情報を読み出し、当該音声合成用情報を 1 以上の音声合成サーバ装置 16 に送信する。音声合成用情報とは、用語の音声合成情報を含む情報であり、音声合成サーバ装置 16 で音声合成のために利用される情報である。音声合成用情報は、2 以上の言語のうちすべての言語または一部の言語についての音声合成情報を含む情報である。音声合成用情報は、音声合成情報と同一の情報でも良いし、音声合成情報に他の情報を加えた情報でも良い。また、音声合成用情報送信部 134 は、音声合成用情報の中で一部に欠けている情報がある用語の音声合成情報は送信しないことは好適である。音声合成用情報送信部 134 は、通常、無線または有線の通信手段で実現されるが、放送手段で実現されても良い。

10

#### 【0052】

表記取得部 135 は、インターネット上の 1 以上のウェブサーバ装置のウェブページから、全言語対辞書格納部 131 に存在しない用語の表記を取得する。表記取得部 135 は、表記を取得する場合に、言語を識別する情報である言語識別子も取得することは好適である。表記取得部 135 は、例えば、用語を取得したウェブページの URL が「.jp」を含む場合は言語識別子「日本語」、「.kr」を含む場合は言語識別子「韓国語」を取得する。また、表記取得部 135 は、例えば、用語を取得した先のデータベースやウェブページの文字コードから、言語を自動識別しても良い。さらに、表記取得部 135 は、ウェブページから用語を取得した後、ユーザに問い合わせ、言語を入力してもらうようにしても良い。

20

#### 【0053】

また、表記取得部 135 は、ウェブページから用語を取得し、当該用語をキーとして、全言語対辞書格納部 131 を検索して、全言語対辞書格納部 131 に用語は存在しないと判断した場合に、用語を取得した、としても良い。つまり、かかる場合、表記取得部 135 は、一度、取得した用語が、全言語対辞書格納部 131 に存在する場合、当該用語を捨てるようにしても良い。

30

#### 【0054】

表記取得部 135 は、通常、MPU やメモリや通信手段等から実現され得る。表記取得部 135 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。表記取得部 135 は、いわゆる検索エンジンを起動する処理を行っても良い。かかる場合、表記取得部 135 は、検索エンジンを有しないと考えても、検索エンジンを有すると考えても良い。

#### 【0055】

表記蓄積部 136 は、表記取得部 135 が取得した用語の表記を、全言語対辞書格納部 131 に蓄積する。通常、表記蓄積部 136 は、表記取得部 135 が取得した言語識別子に対応する言語の表記として、取得した用語の表記を蓄積する。表記蓄積部 136 は、通常、MPU やメモリ等から実現され得る。表記蓄積部 136 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。

40

#### 【0056】

情報受付部 137 は、1 以上の第三端末装置 17 から、用語情報のうちのいずれかの情報を受け付ける。受け付ける情報は、例えば、HMM、表記、読み、アクセント、トーン、フラグ情報などである。情報受付部 137 は、通常、言語と、対応する表記等を識別する情報とともに、情報を受け付ける。つまり、受け付けた情報が、どの言語のどの表記または読み等に対応する情報であるかが判断できる状況であることは言うまでもない。ここ

50

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。情報受付部 137 は、例えば、有線もしくは無線の通信手段で実現され得る。

【0057】

情報蓄積部 138 は、情報受付部 137 が受け付けた情報を、対応する言語の対応する用語の表記に対応付けて蓄積する。また、情報受付部 137 が表記を受け付けた場合、対応する言語の領域に蓄積する。また、情報受付部 137 が表記を受け付け、かつ読みなどの他の情報が既に存在する場合、対応する言語の対応する用語の読み等に対応付けて、表記を蓄積する。情報蓄積部 138 は、通常、MPU やメモリ等から実現され得る。情報蓄積部 138 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。

10

【0058】

出力部 139 は、全言語用語情報または全言語用語情報の一部の情報を出力する。出力部 139 は、通常、第三端末装置 17 からの指示、またはユーザからの指示により、情報を出力する。ここでの出力とは、通常、指示を送信した第三端末装置 17 への、情報の送信である。出力部 139 は、全言語用語情報または全言語用語情報の一部の情報を出力する場合に、2 以上のすべての言語について、予め決められたすべての情報が存在する場合と、予め決められたすべての情報のうちの一部の情報が存在しない場合とにより、視覚的に異なる態様で、全言語用語情報または全言語用語情報の一部の情報を出力することは好適である。予め決められたすべての情報は、全言語対辞書格納部 131 が有する構造情報に対応する情報である。出力部 139 は、通常、無線または有線の通信手段で実現されるが、放送手段で実現されても良い。

20

【0059】

音声認識用情報格納部 141 は、2 以上の言語のうちすべての言語または 2 以上の一部の言語について、音声認識用情報を格納し得る。音声認識用情報格納部 141 は、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。

【0060】

通常、辞書サーバ装置 13 から音声認識用情報を受信することにより、音声認識用情報格納部 141 に音声認識用情報が記憶される。ただし、音声認識用情報格納部 141 に音声認識用情報が記憶される過程は、問わない。例えば、記録媒体を介して音声認識用情報が音声認識用情報格納部 141 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音声認識用情報が音声認識用情報格納部 141 で記憶されるようになってよい。

30

【0061】

音声認識用情報受信部 142 は、辞書サーバ装置 13 から、2 以上の言語のうちすべての言語または 2 以上の一部の言語についての音声認識用情報を受信する。音声認識用情報受信部 142 は、音声認識情報送信指示手段 1452 における辞書サーバ装置 13 への指示の送信に対応して、辞書サーバ装置 13 から音声認識情報を受信しても良い。音声認識用情報受信部 142 は、通常、無線または有線の通信手段で実現されるが、放送を受信する手段で実現されても良い。

40

【0062】

音声認識用情報蓄積部 143 は、音声認識用情報受信部 142 が受信した音声認識用情報を音声認識用情報格納部 141 に蓄積する。また、音声認識用情報蓄積部 143 は、音声認識用情報受信部 142 が受信した音声認識情報を音声認識用情報格納部 141 に蓄積する。音声認識用情報蓄積部 143 は、通常、MPU やメモリ等から実現され得る。音声認識用情報蓄積部 143 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。

50

## 【 0 0 6 3 】

音声情報受信部 1 4 4 は、第一端末装置 1 1 に入力された音声の情報である音声情報を受信する。音声情報受信部 1 4 4 は、第一端末装置 1 1 から直接、または間接的に音声情報を受信する。音声情報受信部 1 4 4 は、通常、無線または有線の通信手段で実現されるが、放送を受信する手段で実現されても良い。

## 【 0 0 6 4 】

音声認識部 1 4 5 は、音声情報受信部 1 4 4 が受信した音声情報を、音声認識用情報格納部 1 4 1 の音声認識用情報を用いて音声認識し、音声認識結果を取得する。音声認識結果は、通常、原言語（第一端末装置 1 1 のユーザ A が発話した音声の言語）の文字列である。音声認識部 1 4 5 が行う音声認識方法は、いかなる音声認識方法でも良い。音声認識部 1 4 5 は公知技術であるので詳細な説明を省略する。音声認識部 1 4 5 は、通常、MPU やメモリ等から実現され得る。音声認識部 1 4 5 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。

10

## 【 0 0 6 5 】

音声認識判断手段 1 4 5 1 は、音声情報受信部 1 4 4 が受信した音声情報に対する音声認識処理が成功であるか失敗であるかを判断する。音声認識判断手段 1 4 5 1 は、例えば、音声情報受信部 1 4 4 が受信した音声情報が有する情報であり、1 以上の用語に対応する情報である部分音声情報に対応する音声認識情報が、音声認識用情報格納部 1 4 1 に存在するか否かを判断する。なお、ここで、部分音声情報は、通常、音声情報の一部であるが、音声情報と同じである場合もあり得る。また、音声認識判断手段 1 4 5 1 は、音声情報受信部 1 4 4 が受信した音声情報に対して、音声認識手段 1 4 5 3 に音声認識処理を行わせ、その結果が成功であるか失敗であるかを判断しても良い。成功である場合は、部分音声情報に対応する音声認識情報が、音声認識用情報格納部 1 4 1 に存在する場合である。また、失敗である場合は、部分音声情報に対応する音声認識情報が、音声認識用情報格納部 1 4 1 に存在しない場合である。なお、音声認識判断手段 1 4 5 1 は、音声認識結果の尤度が予め決められた値より大きい（以上の）場合に音声認識処理が成功であり、予め決められた値以下の（より小さい）場合に音声認識処理が失敗であると判断しても良い。音声認識判断手段 1 4 5 1 は、通常、MPU やメモリ等から実現され得る。音声認識判断手段 1 4 5 1 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。

20

30

## 【 0 0 6 6 】

音声認識情報送信指示手段 1 4 5 2 は、音声情報受信部 1 4 4 が受信した音声情報に対する音声認識処理が失敗であったと判断した場合に、辞書サーバ装置 1 3 に対して、音声認識情報を送信する指示を行う。例えば、音声認識判断手段 1 4 5 1 が音声認識情報は音声認識用情報格納部 1 4 1 に存在しないと判断した場合に、辞書サーバ装置 1 3 に対して、音声認識情報を送信する指示を行う。なお、この指示は、例えば、部分音声情報と言語識別子等を含む。また、この指示は、例えば、部分音声情報から生成された音素列と言語識別子等を含む。音声認識情報送信指示手段 1 4 5 2 は、通常、無線または有線の通信手段で実現されるが、放送手段で実現されても良い。

40

## 【 0 0 6 7 】

音声認識手段 1 4 5 3 は、前記音声情報受信部が受信した音声情報を、前記音声認識用情報格納部の音声認識用情報を用いて音声認識し、音声認識結果を取得する。また、音声認識手段 1 4 5 3 は、音声認識情報送信指示手段 1 4 5 2 における指示の送信に対応して、辞書サーバ装置 1 3 から音声認識用情報受信部 1 4 2 が受信した音声認識情報を用いて音声認識し、音声認識結果を取得する。また、音声認識手段 1 4 5 3 は、音声認識判断手段 1 4 5 1 が音声認識情報は音声認識用情報格納部 1 4 1 に存在すると判断した場合に、音声認識用情報格納部 1 4 1 の音声認識用情報を用いて音声認識し、音声認識結果を取得する。音声認識手段 1 4 5 3 は、通常、MPU やメモリ等から実現され得る。音声認識手段 1 4 5 3 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

50

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。

【0068】

音声認識結果送信部146は、音声認識部145が取得した音声認識結果を送信する。音声認識結果送信部146は、直接的または間接的に、音声認識結果を翻訳サーバ装置15に送信する。音声認識結果送信部146は、通常、無線または有線の通信手段で実現されるが、放送手段で実現されても良い。

【0069】

翻訳用情報格納部151は、2以上の言語のうち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。翻訳用情報は、例えば、翻訳モデルと言語モデルである。翻訳用情報格納部151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。

10

【0070】

通常、辞書サーバ装置13から翻訳用情報を受信することにより、翻訳用情報格納部151に翻訳用情報が記憶される。ただし、翻訳用情報格納部151に翻訳用情報が記憶される過程は問わない。例えば、記録媒体を介して翻訳用情報が翻訳用情報格納部151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翻訳用情報が翻訳用情報格納部151で記憶されるようになってよい。

【0071】

翻訳用情報受信部152は、辞書サーバ装置13から、2以上の言語のうちすべての言語または2以上の一部の言語について、翻訳用情報を受信する。翻訳用情報受信部152は、翻訳情報送信指示手段1552における辞書サーバ装置13への指示の送信に対応して、辞書サーバ装置13から目的言語の用語の表記を受信する。翻訳用情報受信部152は、通常、無線または有線の通信手段で実現されるが、放送を受信する手段で実現されても良い。

20

【0072】

翻訳用情報蓄積部153は、翻訳用情報受信部152が受信した翻訳用情報を翻訳用情報格納部151に蓄積する。また、翻訳用情報蓄積部153は、翻訳用情報受信部152が受信した目的言語の用語の表記を翻訳用情報格納部151に蓄積する。翻訳用情報蓄積部153は、通常、MPUやメモリ等から実現され得る。翻訳用情報蓄積部153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。

30

【0073】

音声認識結果受信部154は、音声認識サーバ装置14が取得した音声認識結果を受信する。音声認識結果受信部154は、音声認識サーバ装置14から直接、または間接的に音声認識結果を受信する。音声認識結果受信部154は、通常、無線または有線の通信手段で実現されるが、放送を受信する手段で実現されても良い。

【0074】

翻訳部155は、音声認識結果受信部154が受信した音声認識結果を、翻訳用情報格納部151の翻訳用情報を用いて、目的言語に翻訳し、翻訳結果を取得する。翻訳部155における翻訳アルゴリズム等と問わない。翻訳部155は公知技術であるので詳細な説明を省略する。翻訳部155は、通常、MPUやメモリ等から実現され得る。翻訳部155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。

40

【0075】

翻訳判断手段1551は、音声認識結果受信部154が受信した音声認識結果に対する翻訳処理が成功であるか失敗であるかを判断する。翻訳判断手段1551は、例えば、音声認識結果受信部154が受信した音声認識結果が有する情報であり、1以上の用語に対応する情報である部分音声認識結果（通常、原言語の用語の文字列）に対応する目的言語の用語の表記が、翻訳用情報格納部151に存在するか否かを判断する。なお、ここで、部分音声認識結果は、通常、音声認識結果の一部であるが、音声認識結果と同じである場

50

合もあり得る。また、翻訳判断手段1551は、音声認識結果受信部154が受信した音声認識結果に対して、翻訳手段1553に翻訳処理を行わせ、その結果が成功であるか失敗であるかを判断しても良い。成功である場合は、部分音声認識結果に対応する目的言語の用語の表記が、翻訳用情報格納部151に存在する場合である。また、失敗である場合は、部分音声認識結果に対応する目的言語の用語の表記が、翻訳用情報格納部151に存在しない場合である。なお、翻訳判断手段1551は、翻訳結果の尤度が予め決められた値より大きい(以上の)場合に翻訳処理が成功であり、予め決められた値以下の(より小さい)場合に翻訳処理が失敗であると判断しても良い。翻訳判断手段1551は、通常、MPUやメモリ等から実現され得る。翻訳判断手段1551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。

10

**【0076】**

翻訳情報送信指示手段1552は、翻訳判断手段1551が目的言語の用語の表記は翻訳用情報格納部151に存在しないと判断した場合に、辞書サーバ装置13に対して、目的言語の用語の表記を送信する指示を行う。この指示は、例えば、原言語の用語の表記と目的言語の言語識別子とを含む。翻訳情報送信指示手段1552は、通常、無線または有線の通信手段で実現されるが、放送手段で実現されても良い。

**【0077】**

翻訳手段1553は、音声認識結果受信部154が受信した音声認識結果を、翻訳用情報格納部151の翻訳用情報を用いて、目的言語に翻訳し、翻訳結果を取得する。また、翻訳手段1553は、翻訳情報送信指示手段1552における指示の送信に対応して、辞書サーバ装置13から翻訳用情報受信部152が受信した目的言語の用語の表記を用いて、音声認識結果を目的言語に翻訳し、翻訳結果を取得する。また、翻訳手段1553は、例えば、翻訳判断手段1551が目的言語の用語の表記は翻訳用情報格納部151に存在すると判断した場合に、翻訳用情報格納部151の目的言語の用語の表記を用いて音声認識結果を翻訳し、翻訳結果を取得する。なお、翻訳結果とは、通常、目的言語の用語の文字列である。翻訳手段1553は、通常、MPUやメモリ等から実現され得る。翻訳手段1553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。

20

**【0078】**

翻訳結果送信部156は、翻訳部155が取得した翻訳結果を送信する。翻訳結果送信部156は、音声合成サーバ装置16に直接的、または間接的に翻訳結果を送信する。翻訳結果送信部156は、通常、無線または有線の通信手段で実現されるが、放送手段で実現されても良い。

30

**【0079】**

音声合成用情報格納部161は、2以上の言語のうちすべての言語または2以上の一部の言語について、音声合成用情報を格納し得る。音声合成用情報格納部161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。

**【0080】**

通常、辞書サーバ装置13から音声合成用情報を受信することにより、音声合成用情報格納部161に音声合成用情報が記憶される。ただし、音声合成用情報格納部161に音声合成用情報が記憶される過程は問わない。例えば、記録媒体を介して音声合成用情報が音声合成用情報格納部161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音声合成用情報が音声合成用情報格納部161で記憶されるようになってよい。

40

**【0081】**

音声合成用情報受信部162は、辞書サーバ装置13から、2以上の言語のうちすべての言語または2以上の一部の言語について、音声合成用情報を受信する。音声合成用情報受信部162は、音声合成情報送信指示手段1652における辞書サーバ装置13への指示の送信に対応して、辞書サーバ装置13から音声合成情報を受信する。音声合成用情

50



報受信部 162 は、通常、無線または有線の通信手段で実現されるが、放送を受信する手段で実現されても良い。

【0082】

音声合成用情報蓄積部 163 は、音声合成用情報受信部 162 が受信した音声合成用情報を音声合成用情報格納部 161 に蓄積する。また、音声合成用情報蓄積部 163 は、音声合成用情報受信部 162 が受信した音声合成情報を音声合成用情報格納部 161 に蓄積する。音声合成用情報蓄積部 163 は、通常、MPU やメモリ等から実現され得る。音声合成用情報蓄積部 163 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。

10

【0083】

翻訳結果受信部 164 は、翻訳サーバ装置 15 が取得した翻訳結果を受信する。翻訳結果受信部 164 は、翻訳サーバ装置 15 から直接、または間接的に翻訳結果を受信する。翻訳結果受信部 164 は、通常、無線または有線の通信手段で実現されるが、放送を受信する手段で実現されても良い。

【0084】

音声合成部 165 は、翻訳結果受信部 164 が受信した翻訳結果を、音声合成用情報格納部 161 の音声合成用情報を用いて音声合成し、音声合成結果を取得する。音声合成のアルゴリズム等と問わない。音声合成部 165 は公知技術であるので詳細な説明を省略する。音声合成部 165 は、通常、MPU やメモリ等から実現され得る。音声合成部 165 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。

20

【0085】

音声合成判断手段 1651 は、翻訳結果受信部 164 が受信した翻訳結果に対する音声合成処理が成功であるか失敗であるかを判断する。音声合成判断手段 1651 は、例えば、翻訳結果受信部 164 が受信した翻訳結果が有する情報であり、1 以上の用語に対応する情報である部分翻訳結果に対応する音声合成情報が、音声合成用情報格納部 161 に存在するか否かを判断する。なお、ここで、部分翻訳結果は、通常、翻訳結果の一部であるが、翻訳結果と同じである場合もあり得る。また、音声合成判断手段 1651 は、翻訳結果受信部 164 が受信した翻訳結果に対して、音声合成手段 1653 に音声合成処理を行わせ、その結果が成功であるか失敗であるかを判断しても良い。成功である場合は、部分翻訳結果に対応する音声合成情報が、音声合成用情報格納部 161 に存在する場合である。また、失敗である場合は、部分翻訳結果に対応する音声合成情報が、音声合成用情報格納部 161 に存在しない場合である。なお、音声合成判断手段 1651 は、音声合成結果の尤度が予め決められた値より大きい（以上の）場合に音声合成処理が成功であり、予め決められた値以下の（より小さい）場合に音声合成処理が失敗であると判断しても良い。音声合成判断手段 1651 は、通常、MPU やメモリ等から実現され得る。音声合成判断手段 1651 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。

30

【0086】

音声合成情報送信指示手段 1652 は、翻訳結果受信部 164 が受信した翻訳結果に対する音声合成処理が失敗であると判断した場合に、辞書サーバ装置 13 に対して、音声合成情報を送信する指示を行う。例えば、音声合成情報送信指示手段 1652 は、音声合成判断手段 1651 が音声合成情報は音声合成用情報格納部 161 に存在しないと判断した場合に、辞書サーバ装置 13 に対して、音声合成情報を送信する指示を行う。なお、この指示は、例えば、目的言語の用語の表記（部分翻訳結果）と目的言語の言語識別子を含む。音声合成情報送信指示手段 1652 は、通常、無線または有線の通信手段で実現されるが、放送手段で実現されても良い。

40

【0087】

音声合成手段 1653 は、翻訳結果受信部 164 が受信した翻訳結果を、音声合成用情

50

報格納部 161 の音声合成用情報を用いて音声合成し、音声合成結果を取得する。また、音声合成手段 1653 は、音声合成情報送信指示手段 1652 における指示の送信に対応して、辞書サーバ装置 13 から音声合成用情報受信部 162 が受信した音声合成情報を用いて部分翻訳結果を音声合成し、音声合成結果を取得する。また、例えば、音声合成手段 1653 は、音声合成判断手段 1651 が音声合成情報は音声合成用情報格納部 161 に存在すると判断した場合に、音声合成用情報格納部 161 の音声合成用情報を用いて部分翻訳結果を音声合成し、音声合成結果を取得しても良い。音声合成結果とは、通常、目的言語の音声の情報である。音声合成手段 1653 は、通常、MPU やメモリ等から実現され得る。音声合成手段 1653 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。

10

## 【0088】

音声合成結果送信部 166 は、音声合成部 165 が取得した音声合成結果を第二端末装置 12 に送信する。音声合成結果送信部 166 は、直接的または間接的に、音声合成結果を第二端末装置 12 に送信する。音声合成結果送信部 166 は、通常、無線または有線の通信手段で実現されるが、放送手段で実現されても良い。

## 【0089】

入力受付部 171 は、ユーザから各種の指示や各種の情報の入力を受け付ける。指示や情報の入力手段は、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。入力受付部 171 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や

20

## 【0090】

情報受信部 172 は、辞書サーバ装置 13 から、情報を受信する。この受信する情報は、全言語用語情報または全言語用語情報の一部の情報である。情報受信部 172 は、通常、無線または有線の通信手段で実現されるが、放送を受信する手段で実現されても良い。

## 【0091】

情報出力部 173 は、情報受信部 172 が受信した情報を出力する。情報出力部 173 は、2 以上のすべての言語について、予め決められたすべての情報が存在する場合と、予め決められたすべての情報のうちの一部の情報が存在しない場合とにより、視覚的に異なる態様で、受信された情報（全言語用語情報または全言語用語情報の一部の情報）を出力する。情報出力部 173 は、ディスプレイやスピーカ等の出力デバイスを含むと考える

30

## 【0092】

入力情報送信部 174 は、入力受付部 171 が受け付けた指示または情報を、辞書サーバ装置 13 に送信する。入力情報送信部 174 は、通常、無線または有線の通信手段で実現されるが、放送手段で実現されても良い。

## 【0093】

次に、音声翻訳システム 1 の動作について説明する。まず、第一端末装置 11 の動作について説明する。第一端末装置 11 の第一音声受付部 111 は、ユーザ A から発声された言語 A の音声を受け付ける。そして、第一音声送信部 112 は、受け付けた音声を音声情報とし、当該音声情報（単に「音声」という場合もある）を送信する。また、第一音声受信部 113 は、ユーザ B が発声した言語 B の音声情報を音声翻訳して得られた言語 A の音声情報を受信する。次に、第一音声出力部 114 は、言語 A の音声情報を、音声として出力する。

40

## 【0094】

次に、第二端末装置 12 の動作について説明する。第二端末装置 12 の第二音声受信部 123 は、ユーザ A から発声された言語 A の音声を音声翻訳して得られた言語 B の音声情報を受信する。そして、第二音声出力部 124 は、第二音声受信部 123 が受信した言語 B の音声情報を、音声として出力する。また、第二音声受付部 121 は、第二端末装置 1

50

2のユーザBから、言語Bの音声を受け付ける。そして、第二音声送信部122は、第二音声受付部121が受け付けた言語Bの音声を音声情報として送信する。

【0095】

次に、辞書サーバ装置13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。

【0096】

(ステップS701) 辞書サーバ装置13は、外部の装置から指示を受信したか否かを判断する。指示を受信すればステップS702に行き、指示を受信しなければステップS706に行く。

【0097】

(ステップS702) 辞書サーバ装置13は、ステップS701で受け付けた指示が情報の送信指示であるか否かを判断する。情報の送信指示であればステップS703に行き、情報の送信指示でなければステップS705に行く。

【0098】

(ステップS703) 辞書サーバ装置13の音声認識用情報送信部132、翻訳用情報送信部133、または音声合成用情報送信部134は、ステップS701で受け付けた指示に対応する情報を、全言語対辞書格納部131から検索し、音声認識のために必要な情報、または翻訳のために必要な情報、または音声合成のために必要な情報を取得する。さらに具体的には、例えば、音声認識用情報送信部132は、ステップS701で受け付けた指示が有する音声認識に失敗した音声情報と言語識別子とをキーとして、全言語対辞書格納部131を検索し、音声認識情報(例えば、音響モデル)を取得する。また、例えば、翻訳用情報送信部133は、ステップS701で受け付けた指示が有する翻訳に失敗した音声認識結果(一部である場合もある)と原言語の識別子と目的言語の識別子とをキーとして、全言語対辞書格納部131を検索し、翻訳情報(例えば、目的言語の用語の表記)を取得する。さらに、例えば、音声合成用情報送信部134は、ステップS701で受け付けた指示が有する音声合成に失敗した翻訳結果(一部である場合もある)と目的言語の識別子とをキーとして、全言語対辞書格納部131を検索し、音声合成情報(例えば、用語の読みとアクセントの情報)を取得する。

【0099】

(ステップS704) 音声認識用情報送信部132、翻訳用情報送信部133、または音声合成用情報送信部134は、ステップS703で取得した情報を送信する。送信先は、指示を送信してきた装置(音声認識サーバ装置14、翻訳サーバ装置15、または音声合成サーバ装置16)である。ステップS701に戻る。

【0100】

(ステップS705) 辞書サーバ装置13は、ステップS701で受け付けた指示に対応する処理を行う。例えば、指示が情報の出力指示(送信指示)である場合、出力部139は、指示に対応する情報を全言語対辞書格納部131から検索し、検索により取得した情報を、指示を送信してきた装置(通常、第三端末装置17)に送信する。ステップS701に戻る。

【0101】

(ステップS706) 情報受付部137は、外部の装置(通常、第三端末装置17)から情報を受信したか否かを判断する。情報を受信すればステップS707に行き、情報を受信しなければステップS708に行く。

【0102】

(ステップS707) 情報蓄積部138は、ステップS706で受信した情報を、全言語対辞書格納部131内の対応する領域に蓄積する。なお、情報とは、例えば、用語の読み、HMM、アクセントの情報、トーンの情報などであり、音声認識、翻訳、または音声合成に利用できる情報である。また、対応する領域とは、以下のことを言う。つまり、受信した情報は、例えば、言語識別子と用語を特定する情報(用語の表記など)と追記する情報(用語の読み、HMM、アクセントの情報、トーンの情報など)が含まれる。そして、情報蓄積部138は、言語識別子と用語を特定する情報を用いて、追記する情報を書き

10

20

30

40

50

込むべき領域を決定し、当該領域に追記する情報を蓄積する。ステップS701に戻る。

【0103】

(ステップS708)音声認識用情報送信部132、翻訳用情報送信部133、または音声合成用情報送信部134は、情報を送信するタイミングであるか否かを判断する。情報を送信するタイミングであればステップS714に行き、情報を送信するタイミングでなければステップS709に行く。ここで、情報とは、音声認識用情報、翻訳用情報、または音声合成用情報である。また、情報を送信するタイミングは、例えば、ユーザからの指示を受け付けたタイミングなどである。さらに、情報の送信先は、音声認識用情報送信部132、翻訳用情報送信部133、または音声合成用情報送信部134が予め格納している。

10

【0104】

(ステップS709)表記取得部135は、ウェブ上のクロールを開始するタイミングであるか否かを判断する。クロールを開始するタイミングであればステップS710に行き、クロールを開始するタイミングでなければステップS701に戻る。なお、例えば、表記取得部135は、定期的にクロールを開始するタイミングであると判断する。

【0105】

(ステップS710)表記取得部135は、ウェブ上をクロールし、用語の表記を取得する。なお、表記取得部135は、更新されたウェブページを検索し、旧ウェブページと比較し、更新箇所を抽出し、当該更新箇所から用語(名詞、名詞句、動詞、形容詞など)を取得することは好適である。なお、用語の表記は、適宜、「用語」とも言う。また、クロールする技術は公知技術であるので詳細な説明を省略する。さらに、表記取得部135は、本ステップで、一の用語を取得しても良いし、複数の用語を取得しても良い。

20

【0106】

(ステップS711)表記取得部135は、ステップS710で取得した1以上の各用語の表記をキーとして、全言語対辞書格納部131を検索する。

【0107】

(ステップS712)表記取得部135は、ステップS711における検索の結果、ステップS710で取得した1以上の各用語の表記が、全言語対辞書格納部131に存在するか否かを判断する。すべての用語の表記が存在すればステップS701に戻り、存在しない用語の表記があればステップS713に行く。

30

【0108】

(ステップS713)表記蓄積部136は、ステップS712で全言語対辞書格納部131に存在しないと判断された1以上の用語の表記を、全言語対辞書格納部131に蓄積する。なお、表記蓄積部136は、用語の表記に対応付けられている言語識別子に対応する領域に、1以上の用語の表記を蓄積する。また、表記取得部135は、用語を取得する際に、用語と対応付けて言語識別子を取得する、とする。ステップS701に戻る。

【0109】

(ステップS714)音声認識用情報送信部132は、音声認識用情報を送信するタイミングであるか否かを判断する。音声認識用情報を送信するタイミングであればステップS715に行き、音声認識用情報を送信するタイミングでなければステップS717に行く。音声認識用情報を送信するタイミングは、例えば、ユーザからの指示を受け付けたタイミングなどである。

40

【0110】

(ステップS715)音声認識用情報送信部132は、全言語対辞書格納部131から音声認識用情報を読み出す。

【0111】

(ステップS716)音声認識用情報送信部132は、ステップS715で読み出した音声認識用情報を、1以上の音声認識サーバ装置14に送信する。なお、1以上の音声認識サーバ装置14に情報を送信するための情報(IPアドレスやURLなど)は、音声認識用情報送信部132が予め保持している、とする。ステップS701に戻る。

50

## 【 0 1 1 2 】

(ステップ S 7 1 7) 翻訳用情報送信部 1 3 3 は、翻訳用情報を送信するタイミングであるか否かを判断する。翻訳用情報を送信するタイミングであればステップ S 7 1 8 に行き、翻訳用情報を送信するタイミングでなければステップ S 7 2 0 に行く。なお、翻訳用情報を送信するタイミングは、例えば、ユーザからの指示を受け付けたタイミングなどである。

## 【 0 1 1 3 】

(ステップ S 7 1 8) 翻訳用情報送信部 1 3 3 は、全言語対辞書格納部 1 3 1 から翻訳用情報を読み出す。

## 【 0 1 1 4 】

(ステップ S 7 1 9) 翻訳用情報送信部 1 3 3 は、ステップ S 7 1 8 で読み出した翻訳用情報を、1 以上の翻訳サーバ装置 1 5 に送信する。なお、1 以上の翻訳サーバ装置 1 5 に情報を送信するための情報 (IP アドレスや URL など) は、翻訳用情報送信部 1 3 3 が予め保持している、とする。ステップ S 7 0 1 に戻る。

## 【 0 1 1 5 】

(ステップ S 7 2 0) 音声合成用情報送信部 1 3 4 は、音声合成用情報を送信するタイミングであるか否かを判断する。音声合成用情報を送信するタイミングであればステップ S 7 2 1 に行き、音声合成用情報を送信するタイミングでなければステップ S 7 0 1 に戻る。なお、音声合成用情報を送信するタイミングは、例えば、ユーザからの指示を受け付けたタイミングなどである。

## 【 0 1 1 6 】

(ステップ S 7 2 1) 音声合成用情報送信部 1 3 4 は、全言語対辞書格納部 1 3 1 から音声合成用情報を読み出す。

## 【 0 1 1 7 】

(ステップ S 7 2 2) 音声合成用情報送信部 1 3 4 は、ステップ S 7 1 8 で読み出した音声合成用情報を、1 以上の音声合成サーバ装置 1 6 に送信する。なお、1 以上の音声合成サーバ装置 1 6 に情報を送信するための情報 (IP アドレスや URL など) は、音声合成用情報送信部 1 3 4 が予め保持している、とする。ステップ S 7 0 1 に戻る。

## 【 0 1 1 8 】

なお、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。

## 【 0 1 1 9 】

次に、音声認識サーバ装置 1 4 の動作について、図 8 のフローチャートを用いて説明する。

## 【 0 1 2 0 】

(ステップ S 8 0 1) 音声情報受信部 1 4 4 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 S 8 0 2 に行き、音声情報を受信しなければステップ S 8 0 9 に行く。

## 【 0 1 2 1 】

(ステップ S 8 0 2) 音声認識手段 1 4 5 3 は、ステップ S 8 0 1 で受信された音声情報に対して、音声認識処理を行い、音声認識結果を得る。

## 【 0 1 2 2 】

(ステップ S 8 0 3) 音声認識判断手段 1 4 5 1 は、ステップ S 8 0 2 における音声認識の結果が成功であるか否かを判断する。ここで、例えば、音声認識結果の尤度を用いて判断する。成功であればステップ S 8 0 4 に行き、成功でなければステップ S 8 0 5 に行く。

## 【 0 1 2 3 】

(ステップ S 8 0 4) 音声認識結果送信部 1 4 6 は、ステップ S 8 0 2 で取得された音声認識結果を、翻訳サーバ装置 1 5 に送信する。ステップ S 8 0 1 に戻る。

## 【 0 1 2 4 】

10

20

30

40

50

(ステップ S 8 0 5) 音声認識情報送信指示手段 1 4 5 2 は、音声情報を音声認識するために必要な情報(音声認識情報)を取得するために、辞書サーバ装置 1 3 に送信する情報を取得する。取得する情報は、例えば、音声認識に失敗した音声情報(一部である場合もある)や、言語を特定する情報(言語識別子)などである。

【0 1 2 5】

(ステップ S 8 0 6) 音声認識情報送信指示手段 1 4 5 2 は、ステップ S 8 0 5 で取得した情報を含む指示を、辞書サーバ装置 1 3 に送信する。この指示は、音声認識情報の送信を促す指示である。

【0 1 2 6】

(ステップ S 8 0 7) 音声認識用情報受信部 1 4 2 は、ステップ S 8 0 6 における指示の送信に対応して、音声認識情報を受信したか否かを判断する。音声認識情報を受信すればステップ S 8 0 8 に行き、音声認識情報を受信しなければステップ S 8 0 7 に戻る。

10

【0 1 2 7】

(ステップ S 8 0 8) 音声認識用情報蓄積部 1 4 3 は、ステップ S 8 0 7 で受信された音声認識情報を音声認識用情報格納部 1 4 1 に蓄積する。

【0 1 2 8】

(ステップ S 8 0 9) 音声認識手段 1 4 5 3 は、ステップ S 8 0 7 で受信された音声認識情報を用いて、ステップ S 8 0 1 で受信された音声情報に対して、音声認識処理を行い、音声認識結果を得る。ステップ S 8 0 4 に行く。

【0 1 2 9】

20

(ステップ S 8 1 0) 音声認識用情報受信部 1 4 2 は、辞書サーバ装置 1 3 から、2 以上の言語のうちすべての言語または 2 以上の一部の言語についての音声認識用情報を受信したか否かを判断する。音声認識用情報を受信すればステップ S 8 1 1 に行き、音声認識用情報を受信しなければステップ S 8 0 1 に戻る。

【0 1 3 0】

(ステップ S 8 1 1) 音声認識用情報蓄積部 1 4 3 は、ステップ S 8 1 0 で受信された音声認識用情報を音声認識用情報格納部 1 4 1 に蓄積する。ステップ S 8 0 1 に戻る。

【0 1 3 1】

なお、図 8 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。

30

【0 1 3 2】

次に、翻訳サーバ装置 1 5 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。

【0 1 3 3】

(ステップ S 9 0 1) 音声認識結果受信部 1 5 4 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 S 9 0 2 に行き、音声認識結果を受信しなければステップ S 9 0 9 に行く。

【0 1 3 4】

(ステップ S 9 0 2) 翻訳手段 1 5 5 3 は、ステップ S 9 0 1 で受信された音声認識結果に対して、翻訳処理を行い、翻訳結果を得る。

【0 1 3 5】

40

(ステップ S 9 0 3) 翻訳判断手段 1 5 5 1 は、ステップ S 9 0 2 における翻訳の結果が成功であるか否かを判断する。ここで、例えば、翻訳結果の尤度を用いて判断する。成功であればステップ S 9 0 4 に行き、成功でなければステップ S 9 0 5 に行く。

【0 1 3 6】

(ステップ S 9 0 4) 翻訳結果送信部 1 5 6 は、ステップ S 9 0 2 で取得された翻訳結果を、音声合成サーバ装置 1 6 に送信する。ステップ S 9 0 1 に戻る。

【0 1 3 7】

(ステップ S 9 0 5) 翻訳情報送信指示手段 1 5 5 2 は、音声認識結果を翻訳するために必要な情報(翻訳情報)を取得するために、辞書サーバ装置 1 3 に送信する情報を取得する。取得する情報は、例えば、翻訳に失敗した音声認識結果(一部である場合もある)

50

と原言語の識別子と目的言語の識別子等である。

【0138】

(ステップS906) 翻訳情報送信指示手段1552は、ステップS905で取得した情報を含む指示を、辞書サーバ装置13に送信する。この指示は、翻訳情報の送信を促す指示である。

【0139】

(ステップS907) 翻訳用情報受信部152は、ステップS906における指示の送信に対応して、翻訳情報を受信したか否かを判断する。翻訳情報を受信すればステップS908に行き、翻訳情報を受信しなければステップS907に戻る。

【0140】

(ステップS908) 翻訳用情報蓄積部153は、ステップS907で受信された翻訳情報を翻訳用情報格納部151に蓄積する。

【0141】

(ステップS909) 翻訳手段1553は、ステップS907で受信された翻訳情報を用いて、ステップS901で受信された音声認識結果に対して、翻訳処理を行い、翻訳結果を得る。ステップS904に行く。

【0142】

(ステップS910) 翻訳用情報受信部152は、辞書サーバ装置13から、2以上の言語のうちすべての言語または2以上の一部の言語についての翻訳用情報を受信したか否かを判断する。翻訳用情報を受信すればステップS911に行き、翻訳用情報を受信しなければステップS901に戻る。

【0143】

(ステップS911) 翻訳用情報蓄積部153は、ステップS910で受信された翻訳用情報を翻訳用情報格納部151に蓄積する。ステップS901に戻る。

【0144】

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。

【0145】

次に、音声合成サーバ装置16の動作について、図10のフローチャートを用いて説明する。

【0146】

(ステップS1001) 翻訳結果受信部164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S1002に行き、翻訳結果を受信しなければステップS1009に行く。

【0147】

(ステップS1002) 音声合成手段1653は、ステップS1001で受信された音声情報に対して、音声合成処理を行い、音声合成結果を得る。

【0148】

(ステップS1003) 音声合成判断手段1651は、ステップS1002における音声合成の結果が成功であるか否かを判断する。ここで、例えば、音声合成結果の尤度を用いて判断する。成功であればステップS1004に行き、成功でなければステップS1005に行く。

【0149】

(ステップS1004) 音声合成結果送信部166は、ステップS1002で取得された音声合成結果を、第二端末装置12に送信する。ステップS1001に戻る。

【0150】

(ステップS1005) 音声合成情報送信指示手段1652は、翻訳結果を音声合成するために必要な情報(音声合成情報)を取得するために、辞書サーバ装置13に送信する情報を取得する。取得する情報は、例えば、音声合成に失敗した翻訳結果(一部である場合もある)と目的言語の識別子である。

10

20

30

40

50

## 【0151】

(ステップS1006)音声合成情報送信指示手段1652は、ステップS1005で取得した情報を含む指示を、辞書サーバ装置13に送信する。この指示は、音声合成情報の送信を促す指示である。

## 【0152】

(ステップS1007)音声合成用情報受信部162は、ステップS1006における指示の送信に対応して、音声合成情報を受信したか否かを判断する。音声合成情報を受信すればステップS1008に行き、音声合成情報を受信しなければステップS1007に戻る。

## 【0153】

(ステップS1008)音声合成用情報蓄積部163は、ステップS1007で受信された音声合成情報を音声合成用情報格納部161に蓄積する。

## 【0154】

(ステップS1009)音声合成手段1653は、ステップS1007で受信された音声合成情報を用いて、ステップS1001で受信された翻訳結果に対して、音声合成処理を行い、音声合成結果を得る。ステップS1004に行く。

## 【0155】

(ステップS1010)音声合成用情報受信部162は、辞書サーバ装置13から、2以上の言語のうちすべての言語または2以上の一部の言語についての音声合成用情報を受信したか否かを判断する。音声合成用情報を受信すればステップS1011に行き、音声合成用情報を受信しなければステップS1001に戻る。

## 【0156】

(ステップS1011)音声合成用情報蓄積部163は、ステップS1010で受信された音声合成用情報を音声合成用情報格納部161に蓄積する。ステップS1001に戻る。

## 【0157】

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。

## 【0158】

次に、第三端末装置17の動作について説明する。第三端末装置17の入力受付部171は、ユーザから各種の指示や各種の情報の入力を受け付ける。例えば、入力受付部171は、ユーザから全言語対辞書の出力指示を受け付ける。すると、入力情報送信部174は、入力受付部171が受け付けた出力指示を、辞書サーバ装置13に送信する。そして、かかる出力指示の送信に対応して、情報受信部172は、辞書サーバ装置13から、全言語対辞書の全部または一部を受信する。次に、情報出力部173は、情報受信部172が受信した全言語対辞書の全部または一部を出力する。なお、かかる出力の際に、情報出力部173は、2以上のすべての言語について、予め決められたすべての情報が存在する場合と、予め決められたすべての情報のうち一部の情報が存在しない場合とにより、視覚的に異なる態様で、受信された情報(全言語用語情報または全言語用語情報の一部の情報)を出力する。つまり、抜けている情報が存在する用語において、情報が抜けていることをユーザに明示するように、全言語対辞書の全部または一部(1以上の用語の情報)を出力する。

## 【0159】

また、入力受付部171は、ユーザから全言語対辞書内の抜けている情報の入力を受け付ける。そして、入力情報送信部174は、入力受付部171が受け付けた情報を、辞書サーバ装置13に送信する。そして、全言語対辞書に、かかる情報が蓄積されることとなる。このことにより、全言語対辞書が充実していく。

## 【0160】

以下、本実施の形態における音声翻訳システム1の具体的な動作について説明する。音声翻訳システム1の概念図は図1である。

10

20

30

40

50



## 【0161】

今、辞書サーバ装置13の全言語対辞書格納部131は、図11の全言語対辞書を格納している。図11の全言語対辞書は、音声翻訳における辞書間の不整合の問題を解決するために、以下のような構造を有する。全言語対辞書は、音声認識辞書、翻訳用対訳辞書、および合成音声用辞書に必要な情報を全言語対で一元管理している情報である。具体的には、全言語対辞書は、構造情報「日本語<表記><読み><音響モデル><アクセント>・・・」「英語<表記><読み><音響モデル><アクセント><トーン>・・・」「中国語<表記><読み><音響モデル><アクセント><声調>・・・」「ドイツ語<表記><読み><音響モデル><アクセント><フラグ情報>・・・」などを有する、とする。そして、全言語対辞書は、構造情報に対応した言語ごとの用語情報を有する。図11において、用語「大阪」の日本語の用語情報は「<表記>大阪<読み>おおさか<音響モデル>・・・<アクセント>4モーラ0型・・・」である。なお、括弧("<">")で囲まれた情報は、辞書の要素(属性)を示す情報であり、括弧("<">")で囲まれた情報群が、ここでの構造情報である。

10

## 【0162】

かかる状況において、まず、ユーザからの指示により、辞書サーバ装置13の音声認識用情報送信部132は、音声認識用情報を送信するタイミングであると、判断した、とする。

## 【0163】

そして、音声認識用情報送信部132は、全言語対辞書格納部131から、図12に示す音声認識用情報(音声認識辞書と同意義である)を読み出す。音声認識辞書は、各用語ごと、各言語ごとに、「表記」「音響モデル」等の情報を有する。

20

## 【0164】

次に、音声認識用情報送信部132は、読み出した音声認識辞書を、1以上の音声認識サーバ装置14に送信する。

## 【0165】

次に、1以上の各音声認識サーバ装置14の音声認識用情報受信部142は、辞書サーバ装置13から、音声認識辞書を受信する。

## 【0166】

そして、音声認識用情報蓄積部143は、受信された音声認識辞書(音声認識用情報)を音声認識用情報格納部141に蓄積する。

30

## 【0167】

また、次に、ユーザからの指示により、辞書サーバ装置13の翻訳用情報送信部133は、翻訳用情報を送信するタイミングであると判断する。

## 【0168】

そして、翻訳用情報送信部133は、全言語対辞書格納部131から、図13に示す翻訳用情報(翻訳辞書と同意義である)を読み出す。翻訳辞書は、各用語ごと、各言語ごとに、「表記」等の情報を有する。

## 【0169】

次に、翻訳用情報送信部133は、読み出した翻訳辞書を、1以上の翻訳サーバ装置15に送信する。

40

## 【0170】

次に、1以上の各翻訳サーバ装置15の翻訳用情報受信部152は、辞書サーバ装置13から、翻訳辞書を受信する。

## 【0171】

そして、翻訳用情報蓄積部153は、受信された翻訳辞書(翻訳用情報)を翻訳用情報格納部151に蓄積する。

## 【0172】

さらに、音声合成用情報送信部134は、音声合成用情報を送信するタイミングであると判断する。

50

## 【0173】

次に、音声合成用情報送信部134は、全言語対辞書格納部131から、図14に示す音声合成用情報（音声合成辞書と同意義である）を読み出す。音声合成辞書は、各用語ごと、各言語ごとに、「表記」「読み」「アクセント」等の情報を有する。

## 【0174】

次に、音声合成用情報送信部134は読み出した音声合成辞書を、1以上の音声合成サーバ装置16に送信する。

## 【0175】

そして、音声合成用情報受信部162は、辞書サーバ装置13から、音声合成辞書を受信する。

## 【0176】

そして、音声合成用情報蓄積部163は、は、受信された音声合成辞書（音声合成用情報）を音声合成用情報格納部161に蓄積する。

## 【0177】

以上の処理により、音声認識サーバ装置14、翻訳サーバ装置15、および音声合成サーバ装置16には、整合性のある辞書が格納され、通常、問題なく音声翻訳ができることとなった。ただし、音声認識サーバ装置14、翻訳サーバ装置15、および音声合成サーバ装置16の各辞書は、それぞれ独自に拡張され得ることもあり、その結果、辞書間の不整合が生じる場合もあり得る。

## 【0178】

次に、音声認識サーバ装置14、翻訳サーバ装置15、および音声合成サーバ装置16に、音声認識、翻訳、および音声合成に必要な辞書が構築された状況において、第一端末装置11のユーザAと、第二端末装置12のユーザBが、音声翻訳システムを利用して、電話で会話を行う、とする。なお、ユーザAの言語は、例えば、日本語である。また、ユーザBの言語は、例えば、英語である。

## 【0179】

そして、第一端末装置11は、図15に示す音声翻訳制御情報を保持している、とする。さらに、第二端末装置12は、図16に示す音声翻訳制御情報を保持している、とする。音声翻訳制御情報は、ユーザが発声する言語である原言語の情報（ここでは<原言語>の情報）、対話の相手が発声する言語である目的言語の情報（ここでは<目的言語>の情報）、音声認識サーバ装置14と通信するための情報（ここでは<音声認識サーバ>の情報）、翻訳サーバ装置15と通信するための情報（ここでは<翻訳サーバ>の情報）、音声合成サーバ装置16と通信するための情報（ここでは<音声合成サーバ>の情報）、第二端末装置12または第一端末装置11の識別子（ここでは<相手端末>の情報）、および第一端末装置11または第二端末装置12の識別子（ここでは<自端末>の情報）などを有する。なお、<音声認識サーバ>、<翻訳サーバ>、<音声合成サーバ>の各情報は、ここでは、各装置のIPアドレスであるが、URLや電話番号など、他の情報でも良いことは言うまでもない。また、<相手端末>、<自端末>の各情報は、ここでは、電話番号であるが、IPアドレスやMACアドレスなど他の情報でも良いことは言うまでもない。

## 【0180】

そして、会話中において、ユーザAが第一端末装置11に対して「大阪」と発声した、とする。

## 【0181】

次に、第一端末装置11は、「大阪」の音声情報を構成する。そして、第一端末装置11は、図15に示す音声翻訳制御情報を読み出し、「<音声認識サーバ>186.221.1.27」で特定される音声認識サーバ装置14に、「大阪」の音声情報と、図15に示す音声翻訳制御情報を送信する。

## 【0182】

次に、音声認識サーバ装置14は、「大阪」の音声情報と、図15に示す音声翻訳制御

10

20

30

40

50

情報を受信する。

【0183】

次に、音声認識サーバ装置14の音声認識手段1453は、図15に示す音声翻訳制御情報の「<原言語>日本語」を取得する。そして、音声認識手段1453は、日本語の音響モデル(図12参照)を用いて、受信された「大阪」の音声情報に対して、音声認識処理を行い、音声認識結果(文字列「大阪」を有する)を得る。

【0184】

次に、音声認識判断手段1451は、音声認識結果が有する尤度が予め決められた閾値以上であるか否か(音声認識が成功であるか否か)を判断する。ここで、音声認識が成功であった、とする。

【0185】

次に、音声認識結果送信部146は、取得された音声認識結果を、図15に示す音声翻訳制御情報が有する「<翻訳サーバ>225.68.21.129」で示される翻訳サーバ装置15に送信する。また、音声認識結果送信部146は、図15に示す音声翻訳制御情報も、「<翻訳サーバ>225.68.21.129」で示される翻訳サーバ装置15に送信する。

【0186】

次に、翻訳サーバ装置15の音声認識結果受信部154は、音声認識結果(「大阪」を有する)および音声翻訳制御情報を受信する。

【0187】

次に、翻訳手段1553は、音声翻訳制御情報が有する「<原言語>日本語」および「<目的言語>英語」を読み出す。

【0188】

そして、翻訳手段1553は、音声認識結果(「大阪」を有する)を日本語であると判断し、用語「大阪」と対になる「<目的言語>英語」の用語「Osaka」を、図13の翻訳辞書から読み出す。そして、翻訳手段1553は、翻訳結果「Osaka」を得る。

【0189】

次に、翻訳判断手段1551は、「<目的言語>英語」に対応する用語を翻訳辞書から検索できたので、翻訳結果が成功であると判断する。

【0190】

次に、翻訳結果送信部156は、取得された翻訳結果「Osaka」と図15に示す音声翻訳制御情報とを、「<音声合成サーバ>56.72.128.202」で示される音声合成サーバ装置16に送信する。

【0191】

次に、音声合成サーバ装置16の翻訳結果受信部164は、翻訳結果「Osaka」と図15に示す音声翻訳制御情報とを受信する。

【0192】

そして、音声合成手段1653は、音声翻訳制御情報から「<目的言語>英語」を読み出す。

【0193】

次に、音声合成手段1653は、用語「Osaka」、および「<目的言語>英語」に対応する音声合成情報(読みおよびアクセントなど)を、図14の音声合成辞書から読み出す。

【0194】

次に、音声合成手段1653は、読み出した音声合成情報を用いて、音声合成処理を行い、音声合成結果を得る。

【0195】

次に、音声合成判断手段1651は、音声合成結果の尤度が予め決められた閾値以上であるので、音声合成の結果が成功であると判断する、とする。

【0196】

10

20

30

40

50

次に、音声合成結果送信部 166 は、取得された音声合成結果とを、「<相手端末> 090-1445-1122」で特定される第二端末装置 12 に送信する。

【0197】

次に、第二端末装置 12 の第二音声受信部 123 は、ユーザ A から発声された日本語の音声「大阪」を音声翻訳して得られた英語の音声（「Osaka」の音声）を受信する。

【0198】

そして、第二音声出力部 124 は、第二音声受信部 123 が受信した音声（「Osaka」の音声）を出力する。

【0199】

以上の処理により、音声翻訳が実現する。また、ユーザ B が発声した音声ユーザ A に、翻訳されて届くためには、図 16 の音声翻訳制御情報を利用する。また、音声翻訳の手順は、上記と同様である。

【0200】

また、音声認識、翻訳、音声合成の各処理段階で、各処理が失敗した場合、音声認識サーバ装置 14、翻訳サーバ装置 15、または音声合成サーバ装置 16 は、リアルタイムに、辞書サーバ装置 13 に問い合わせ、必要な情報を送信してもらう。そして、音声認識サーバ装置 14、翻訳サーバ装置 15、または音声合成サーバ装置 16 は、各処理に必要な情報を辞書サーバ装置 13 から受信した後、音声認識、翻訳、または音声合成の処理を進めることは上述した通りである。

【0201】

さらに、辞書サーバ装置 13 の表記取得部 135 は、ウェブ上のクローリングを開始するタイミングであると判断した、とする。

【0202】

そして、表記取得部 135 は、ウェブ上をクローリングし、例えば、用語「名古屋」の表記「名古屋」を取得した、とする。そして、表記取得部 135 は、取得した用語の表記「名古屋」をキーとして、全言語対辞書を検索する。そして、表記取得部 135 は、検索の結果、取得した用語の表記「名古屋」が、全言語対辞書に存在しないと判断した、とする。

【0203】

次に、表記蓄積部 136 は、全言語対辞書に存在しないと判断された用語の表記「名古屋」を、全言語対辞書に蓄積する。そして、全言語対辞書は、例えば、図 17 のようになる。図 17 において、用語「名古屋」について、表記のみ存在し、読みや音響モデルが存在していない状況です。

【0204】

かかる状況において、第三端末装置 17 の入力受付部 171 は、ユーザから全言語対辞書の出力指示を受け付けた、とする。

【0205】

次に、入力情報送信部 174 は、入力受付部 171 が受け付けた出力指示を、辞書サーバ装置 13 に送信する。

【0206】

次に、辞書サーバ装置 13 は、出力指示を受信する。そして、出力部 139 は、出力指示に対応する情報を全言語対辞書から検索し、検索により取得した情報を、指示を送信してきた装置（通常、第三端末装置 17）に送信する。

【0207】

次に、そして、出力指示の送信に対応して、第三端末装置 17 の情報受信部 172 は、辞書サーバ装置 13 から、全言語対辞書の一部を受信する。

【0208】

次に、次に、情報出力部 173 は、情報受信部 172 が受信した全言語対辞書の一部を出力する。かかる出力例は、図 18 である。図 18 において、2 以上のすべての言語について、予め決められたすべての情報が存在する場合（用語「大阪」）と、予め決められたすべての情報のうちの一部の情報が存在しない場合（用語「名古屋」）とにより、視覚的

10

20

30

40

50

に異なる態様で情報を出力している。図 18 において、空のデータ領域（セル）は、網掛けになっている。これは、不足している情報の入力を、第三端末装置 17 のユーザに促すためである。

【0209】

そして、ユーザは、例えば、用語「名古屋」の読み「なごや」や日本語の音響モデルや、韓国語の表記や読み等、英語の表記や読み等を入力する。次に、入力受付部 171 は、ユーザからの入力を受け付ける。

【0210】

次に、第三端末装置 17 の入力情報送信部 174 は、入力受付部 171 が受け付けた情報を、辞書サーバ装置 13 に送信する。

10

【0211】

次に、辞書サーバ装置 13 の情報受付部 137 は、第三端末装置 17 から、用語情報のうちのいくつかの情報（用語「名古屋」の読み「なごや」や日本語の音響モデルや、韓国語の表記や読み等、英語の表記や読み等）を受け付ける。

【0212】

そして、情報蓄積部 138 は、情報受付部 137 が受け付けた情報を、対応する言語の対応する用語の表記に対応付けて蓄積する。

【0213】

以上の処理により、辞書サーバ装置 13 の全言語対辞書は充実していく。

【0214】

以上、本実施の形態によれば、音声翻訳において使用する音声認識辞書、翻訳辞書、および音声合成辞書において、ある辞書に存在する用語の情報が他の辞書には存在しない等という辞書の間での不整合を回避できる。その結果、音声翻訳が正常に行われるようになる。

20

【0215】

また、本実施の形態によれば、音声認識、翻訳、音声合成の各処理段階で、各処理が失敗する場合、音声認識サーバ装置 14、翻訳サーバ装置 15、または音声合成サーバ装置 16 は、リアルタイムに、辞書サーバ装置 13 に問い合わせ、必要な情報を送信してもらえる。その結果、極めて高い確率で、音声翻訳が行われる。

【0216】

さらに、本実施の形態によれば、ウェブ上をクロールしたり、ユーザからの情報の入力を受け付けたりすることにより、辞書サーバ装置 13 の全言語対辞書を充実させることができるようになる。このため、新出の用語などを含む、種々の用語に対して、音声翻訳が可能となり得る。

30

【0217】

なお、本実施の形態において、主として、音声認識部 145 は、音声認識判断手段 1451、音声認識情報送信指示手段 1452、音声認識手段 1453 を具備するものとして説明した。ただし、音声認識部 145 は、音声認識判断手段 1451、および音声認識情報送信指示手段 1452 を具備しなくても良い。かかる場合、音声認識部 145 は、音声認識処理を失敗しない、という前提である。また、本実施の形態において、主として、翻訳部 155 は、翻訳判断手段 1551、翻訳情報送信指示手段 1552、翻訳手段 1553 を具備するものとして説明した。ただし、翻訳部 155 は、翻訳判断手段 1551、および翻訳情報送信指示手段 1552 を具備しなくても良い。かかる場合、翻訳部 155 は、翻訳処理を失敗しない、という前提である。さらに、音声合成部 165 は、音声合成判断手段 1651、音声合成情報送信指示手段 1652、音声合成手段 1653 を具備するものとして説明した。ただし、音声合成部 165 は、音声合成判断手段 1651、および音声合成情報送信指示手段 1652 を具備しなくても良い。かかる場合、音声合成部 165 は、音声合成処理を失敗しない、という前提である。かかる場合、本実施の形態における音声翻訳システム 1 は、辞書サーバ装置と、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

40

50

辞書サーバ装置は、2以上の言語の各言語について、1以上の単語を有する用語の表記と、用語を音声認識するための情報である音声認識情報と、用語を音声合成するための情報である音声合成情報とを有する用語情報を、前記2以上の言語のすべてについて対応付けて有する全言語用語情報を、2以上格納し得る全言語対辞書格納部と、前記2以上の言語のうちすべての言語または2以上の一部の言語について、前記用語の音声認識情報を含む音声認識用情報を、前記全言語対辞書格納部から取得し、前記1以上の音声認識サーバ装置に送信する音声認識用情報送信部と、前記2以上の言語のうちすべての言語または2以上の一部の言語について、前記用語の表記を含む翻訳用情報を、前記全言語対辞書格納部から取得し、前記1以上の翻訳サーバ装置に送信する翻訳用情報送信部と、前記2以上の言語のうちすべての言語または2以上の一部の言語について、前記用語の音声合成情報を含む音声合成用情報を、前記全言語対辞書格納部から取得し、前記1以上の音声合成サーバ装置に送信する音声合成用情報送信部とを具備し、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、音声認識用情報を格納し得る音声認識用情報格納部と、前記辞書サーバ装置から、前記2以上の言語のうちすべての言語または2以上の一部の言語についての音声認識用情報を受信する音声認識用情報受信部と、前記音声認識用情報受信部が受信した音声認識用情報を前記音声認識用情報格納部に蓄積する音声認識用情報蓄積部と、第一端末装置に入力された音声の情報である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、前記音声認識用情報格納部の音声認識用情報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用情報格納部と、前記辞書サーバ装置から、前記2以上の言語のうちすべての言語または2以上の一部の言語について、音翻訳用情報を受信する翻訳用情報受信部と、前記翻訳用情報受信部が受信した翻訳用情報を前記翻訳用情報格納部に蓄積する翻訳用情報蓄積部と、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、前記翻訳用情報格納部の翻訳用情報を用いて、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、音声合成用情報を格納し得る音声合成用情報格納部と、前記辞書サーバ装置から、前記2以上の言語のうちすべての言語または2以上の一部の言語について、音声合成用情報を受信する音声合成用情報受信部と、前記音声合成用情報受信部が受信した音声合成用情報を前記音声合成用情報格納部に蓄積する音声合成用情報蓄積部と、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、前記音声合成用情報格納部の音声合成用情報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、である。

#### 【0218】

また、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辞書サーバ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、記憶媒体に、2以上の言語の各言語について、1以上の単語を有する用語の表記と、用語を音声認識するための情報である音声認識情報と、用語を音声合成するための情報である音声合成情報とを有する用語情報を、前記2以上の言語のすべてについて対応付けて有する全言語用語情報を、2以上格納しており、コンピュータを、前記2以上の言語のうちすべての言語または2以上の一部の言語について、前記用語の音声認識情報を含む音声認識用情報を、前記記憶媒体から取得し、前記1以上の音声認識サーバ装置に送信する音声認識用情報送信部と、前記2以上の言語のうちすべての言語または2以上の一部の言語について、前記用語の表記を含む翻訳用情報を、前記記憶媒体から取得し、前記1

以上の翻訳サーバ装置に送信する翻訳用情報送信部と、前記2以上の言語のうちすべての言語または2以上の一部の言語について、前記用語の音声合成情報を含む音声合成用情報を、前記記憶媒体から取得し、前記1以上の音声合成サーバ装置に送信する音声合成用情報送信部として機能させるためのプログラム、である。

【0219】

また、上記プログラムにおいて、コンピュータを、インターネット上の1以上のウェブサーバ装置のウェブページから、前記記憶媒体に存在しない用語の表記を取得する表記取得部と、前記表記取得部が取得した用語の表記を、前記記憶媒体に蓄積する表記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ことは好適である。

【0220】

また、上記プログラムにおいて、コンピュータを、1以上の第三端末装置から、用語情報のうちのいずれかの情報を受け付ける情報受付部と、前記情報受付部が受け付けた情報を、対応する言語の対応する用語の表記に対応付けて、前記記憶媒体に蓄積する情報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ことは好適である。

【0221】

また、上記プログラムにおいて、コンピュータを、前記全言語用語情報または前記全言語用語情報の一部の情報を出力する出力部をさらに機能させ、前記出力部は、前記全言語用語情報または前記全言語用語情報の一部の情報を出力する場合に、前記2以上のすべての言語について、予め決められたすべての情報が存在する場合と、予め決められたすべての情報のうちの一部の情報が存在しない場合とにより、視覚的に異なる態様で、前記全言語用語情報または前記全言語用語情報の一部の情報を出力するものとして機能させるためのプログラムであることは好適である。

【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、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、記憶媒体に、2以上の言語のうちすべての言語または2以上の一部の言語について、音声認識用情報を格納しており、コンピュータを、辞書サーバ装置から、前記2以上の言語のうちすべての言語または2以上の一部の言語についての音声認識用情報を受信する音声認識用情報受信部と、前記音声認識用情報受信部が受信した音声認識用情報を前記記憶媒体に蓄積する音声認識用情報蓄積部と、第一端末装置に入力された音声の情報である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、前記記憶媒体の音声認識用情報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。

【0223】

また、上記プログラムにおいて、前記音声認識部は、前記音声情報受信部が受信した音声情報に対する音声認識処理が成功であるか失敗であるかを判断する音声認識判断手段と、前記音声認識判断手段が、前記音声認識処理が失敗であると判断した場合に、前記辞書サーバ装置に対して、音声認識情報を送信する指示を行う音声認識情報送信指示手段と、前記音声情報受信部が受信した音声情報を、前記記憶媒体の音声認識用情報を用いて音声認識し、音声認識結果を取得し、かつ、前記指示の送信に対応して、前記辞書サーバ装置から前記音声認識用情報受信部が受信した音声認識情報を用いて音声認識し、音声認識結果を取得する音声認識手段とを具備し、前記音声認識用情報受信部は、前記指示の送信に対応して、前記辞書サーバ装置から前記音声認識情報を受信するものとして機能させるためのプログラムであることは好適である。

【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、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、記憶媒体に、2以上の言語のうち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ており、コンピュータを、辞書サーバ装置から、2以上の言語のうちすべての言語または2以上の一部の言語について、音翻訳用情報を受信する翻訳用情報受信部と、前記翻訳用情報受信部が受信し

10

20

30

40

50

た翻訳用情報を前記記憶媒体に蓄積する翻訳用情報蓄積部と、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、前記記憶媒体の翻訳用情報を用いて、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。

【0225】

また、上記プログラムにおいて、前記翻訳部は、前記音声認識結果受信部が受信した音声認識結果に対する翻訳処理が成功であるか失敗であるかを判断する翻訳判断手段と、前記翻訳判断手段が、前記翻訳処理が失敗であると判断した場合に、前記辞書サーバ装置に対して、前記目的言語の用語の表記を送信する指示を行う翻訳情報送信指示手段と、前記音声認識結果受信部が受信した音声認識結果を、前記記憶媒体の翻訳用情報を用いて、目的言語に翻訳し、翻訳結果を取得し、かつ、前記指示の送信に対応して、前記辞書サーバ装置から前記翻訳用情報受信部が受信した目的言語の用語の表記を用いて、前記音声認識結果を目的言語に翻訳し、翻訳結果を取得する翻訳手段とを具備し、前記翻訳用情報受信部は、前記指示の送信に対応して、前記辞書サーバ装置から前記目的言語の用語の表記を受信するものとして機能させるためのプログラムであることは好適である。

10

【0226】

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、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、記憶媒体に、2以上の言語のうちすべての言語または2以上の一部の言語について、音声合成用情報を格納しており、コンピュータを、辞書サーバ装置から、2以上の言語のうちすべての言語または2以上の一部の言語について、音声合成用情報を受信する音声合成用情報受信部と、前記音声合成用情報受信部が受信した音声合成用情報を前記記憶媒体に蓄積する音声合成用情報蓄積部と、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、前記記憶媒体の音声合成用情報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。

20

【0227】

また、上記プログラムにおいて、前記音声合成部は、前記翻訳結果受信部が受信した翻訳結果に対する音声合成処理が成功であるか失敗であるかを判断する音声合成判断手段と、前記音声合成判断手段が、前記音声合成処理が失敗であると判断した場合に、前記辞書サーバ装置に対して、音声合成情報を送信する指示を行う音声合成情報送信指示手段と、前記翻訳結果受信部が受信した翻訳結果を、前記記憶媒体の音声合成用情報を用いて音声合成し、音声合成結果を取得し、かつ、前記指示の送信に対応して、前記辞書サーバ装置から前記音声合成用情報受信部が受信した音声合成情報を用いて前記翻訳結果を音声合成し、音声合成結果を取得する音声合成手段とを具備し、前記音声合成用情報受信部は、前記指示の送信に対応して、前記辞書サーバ装置から前記音声合成情報を受信するものとして機能させるためのプログラムであることは好適である。

30

【0228】

また、図19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19は、このコンピュータシステム340の概観図であり、図20は、コンピュータシステム340のブロック図である。

40

【0229】

図19において、コンピュータシステム340は、FDドライブ、CD-ROMドライブ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。

【0230】

図20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

50



ブ 3 4 1 1 に接続されたバス 3 4 1 4 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 ROM 3 4 1 5 とに接続され、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 RAM 3 4 1 6 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 3 4 1 7 とを含む。ここでは、図示しないが、コンピュータ 3 4 1 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。

【 0 2 3 1 】

コンピュータシステム 3 4 0 に、上述した実施の形態の音声認識等の機能を実行させるプログラムは、CD-ROM 3 5 0 1、または FD 3 5 0 2 に記憶されて、CD-ROM ドライブ 3 4 1 2 または FD ドライブ 3 4 1 1 に挿入され、さらにハードディスク 3 4 1 7 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 3 4 1 に送信され、ハードディスク 3 4 1 7 に記憶されても良い。プログラムは実行の際に RAM 3 4 1 6 にロードされる。プログラムは、CD-ROM 3 5 0 1、FD 3 5 0 2 またはネットワークから直接、ロードされても良い。

10

【 0 2 3 2 】

プログラムは、コンピュータ 3 4 1 に、上述した実施の形態の音声認識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 3 4 0 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。

20

【 0 2 3 3 】

なお、上記プログラムにおいて、情報を送信する送信ステップや、情報を受信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。

【 0 2 3 4 】

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。

【 0 2 3 5 】

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 (音声情報受信部、音声認識用情報受信部など) 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。

30

【 0 2 3 6 】

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。

【 0 2 3 7 】

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 】

40

【 0 2 3 8 】

以上のように、本発明にかかる音声翻訳システムは、音声翻訳において使用する辞書の間の不整合をなくすることができる、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。

【 符号の説明 】

【 0 2 3 9 】

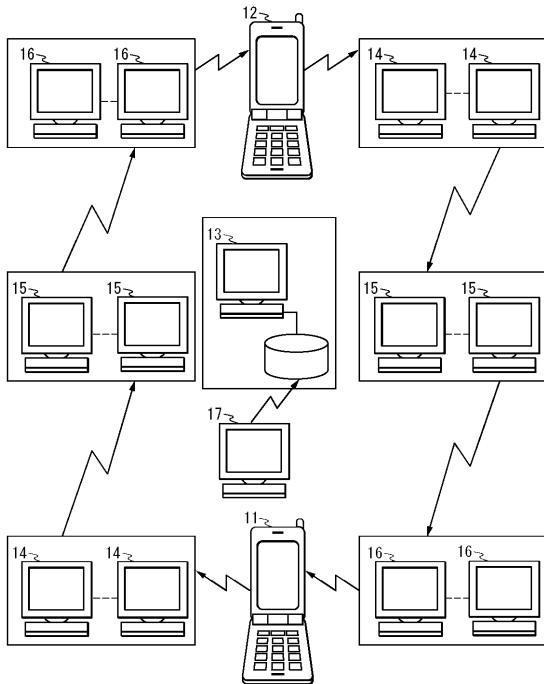
- 1 音声翻訳システム
- 1 1 第一端末装置
- 1 2 第二端末装置
- 1 3 辞書サーバ装置

50

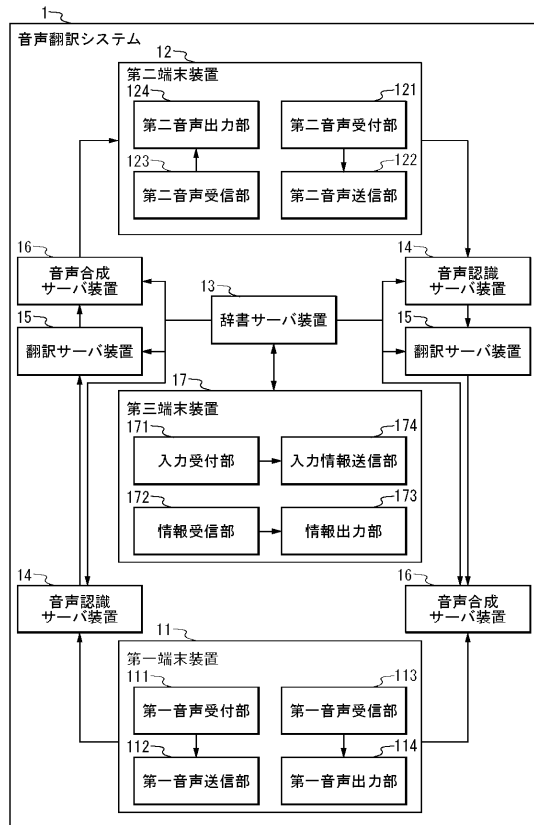
1 4	音声認識サーバ装置	
1 5	翻訳サーバ装置	
1 6	音声合成サーバ装置	
1 7	第三端末装置	
1 1 1	第一音声受付部	
1 1 2	第一音声送信部	
1 1 3	第一音声受信部	
1 1 4	第一音声出力部	
1 2 1	第二音声受付部	
1 2 2	第二音声送信部	10
1 2 3	第二音声受信部	
1 2 4	第二音声出力部	
1 3 1	全言語対辞書格納部	
1 3 2	音声認識用情報送信部	
1 3 3	翻訳用情報送信部	
1 3 4	音声合成用情報送信部	
1 3 5	表記取得部	
1 3 6	表記蓄積部	
1 3 7	情報受付部	
1 3 8	情報蓄積部	20
1 3 9	出力部	
1 4 1	音声認識用情報格納部	
1 4 2	音声認識用情報受信部	
1 4 3	音声認識用情報蓄積部	
1 4 4	音声情報受信部	
1 4 5	音声認識部	
1 4 6	音声認識結果送信部	
1 5 1	翻訳用情報格納部	
1 5 2	翻訳用情報受信部	
1 5 3	翻訳用情報蓄積部	30
1 5 4	音声認識結果受信部	
1 5 5	翻訳部	
1 5 6	翻訳結果送信部	
1 6 1	音声合成用情報格納部	
1 6 2	音声合成用情報受信部	
1 6 3	音声合成用情報蓄積部	
1 6 4	翻訳結果受信部	
1 6 5	音声合成部	
1 6 6	音声合成結果送信部	
1 7 1	入力受付部	40
1 7 2	情報受信部	
1 7 3	情報出力部	
1 7 4	入力情報送信部	
1 4 5 1	音声認識判断手段	
1 4 5 2	音声認識情報送信指示手段	
1 4 5 3	音声認識手段	
1 5 5 1	翻訳判断手段	
1 5 5 2	翻訳情報送信指示手段	
1 5 5 3	翻訳手段	
1 6 5 1	音声合成判断手段	50

- 1 6 5 2 音声合成情報送信指示手段
- 1 6 5 3 音声合成手段

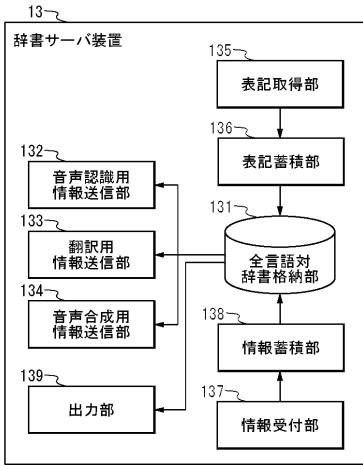
【 図 1 】



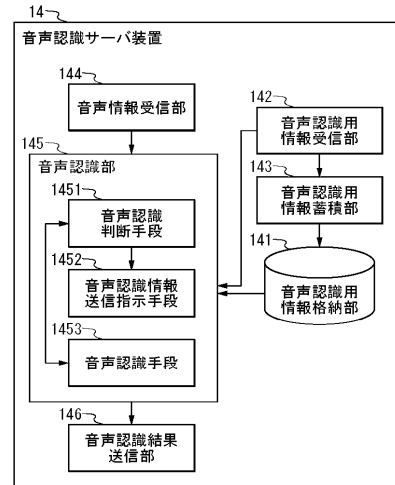
【 図 2 】



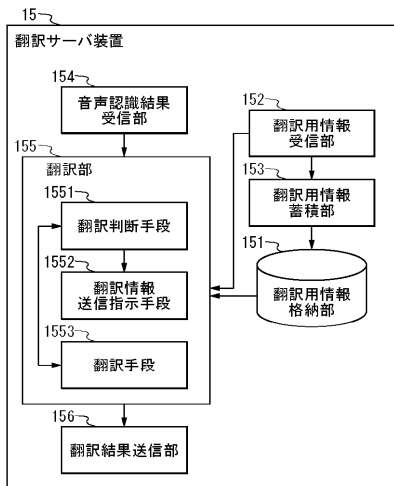
【図3】



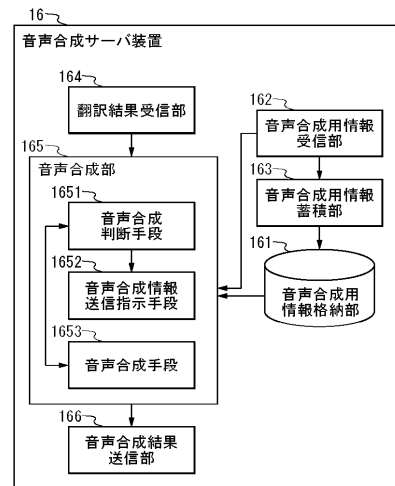
【図4】



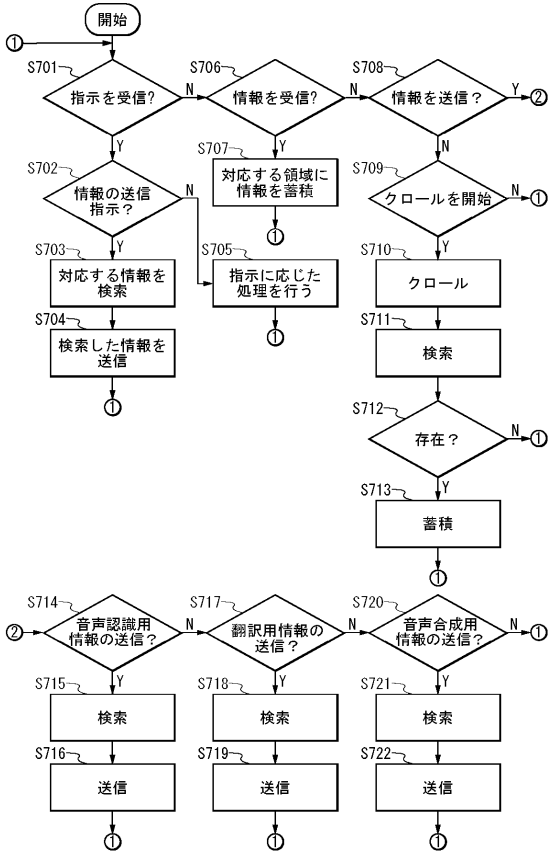
【図5】



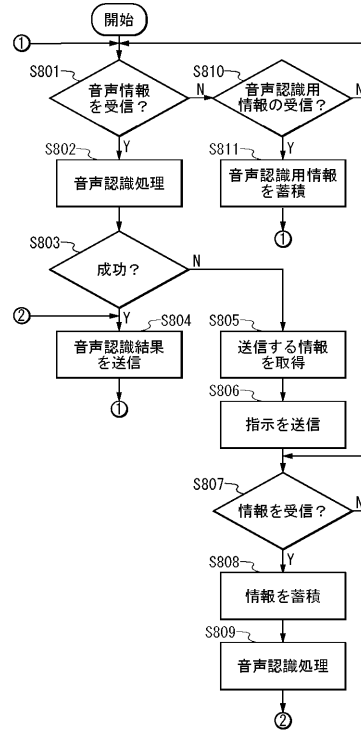
【図6】



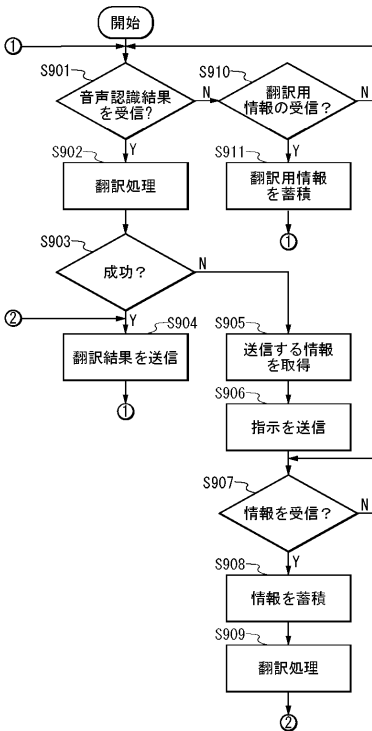
【 図 7 】



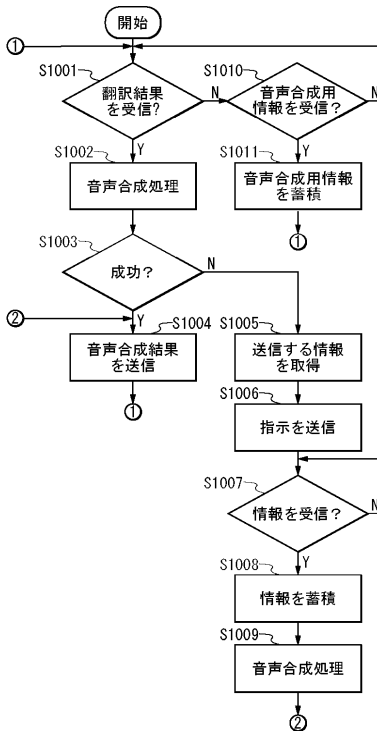
【 図 8 】




【 図 9 】




【 図 10 】



【 図 1 1 】

用語 (辞書エントリ)	項目	言語				---
		日本語	韓国語	中国語	英語	
大阪	表記	大阪	오사카	大阪	Osaka	--
	読み	おおさか	Osaka	Daban	ousa:ka:	--
	音響モデル		----	----	----	--
	アクセント	4t-70型	----	Da4ban3	óusaka	--
	他	----	----	<声調> 四声三声	----	--
東京	表記	東京	도쿄	東京	Tokyo	--
	読み	とうきょう	----	Tong1jing1	----	--
	音響モデル	----	----	----	----	--
	アクセント	----	----	----	tóukyo	--
	他	----	----	----	----	--

【 図 1 2 】

用語 (辞書エントリ)	項目	言語				---
		日本語	韓国語	中国語	英語	
大阪	表記	大阪	오사카	大阪	Osaka	---
	音響モデル		----	----	----	---
	他	----	----	----	----	---
東京	表記	東京	도쿄	東京	Tokyo	---
	音響モデル	----	----	----	----	---
	他	----	----	----	----	---

【 図 1 3 】

用語 (辞書エントリ)	項目	言語				---
		日本語	韓国語	中国語	英語	
大阪	表記	大阪	오사카	大阪	Osaka	---
	他	----	----	----	----	---
東京	表記	東京	도쿄	東京	Tokyo	---
	他	----	----	----	----	---
---	---	---	---	---	---	---

【 図 1 4 】

用語 (辞書エントリ)	項目	言語				---
		日本語	韓国語	中国語	英語	
大阪	表記	大阪	오사카	大阪	Osaka	---
	読み	おおさか	Osaka	Daban	ousa:ka:	---
	アクセント	4t-70型	----	Da4ban3	óusaka	---
	他	----	----	<声調> 四声三声	----	---
	表記	東京	도쿄	東京	Tokyo	---
東京	読み	とうきょう	----	Tong1jing1	----	---
	アクセント	----	----	----	tóukyo	---
	他	----	----	----	----	---
---	---	---	---	---	---	---

【 図 1 5 】

<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>	56.72.128.202
<相手端末>	092-1445-1122
<自端末>	080-1111-2256
...	...

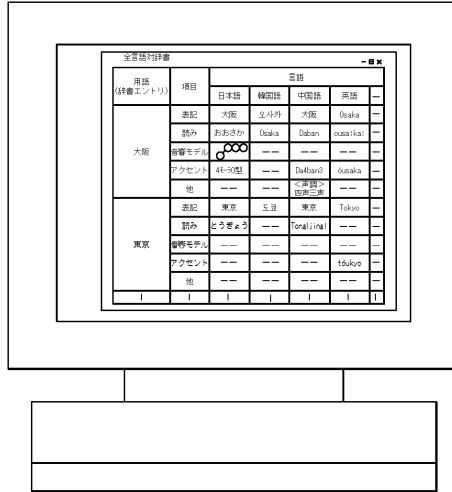
【図16】

<原言語>	英語
<目的言語>	日本語
<音声認識サーバ>	215.76.2.11
<翻訳サーバ>	225.68.21.129
<音声合成サーバ>	88.125.6.218
<相手端末>	080-1111-2256
<自端末>	090-1645-1122

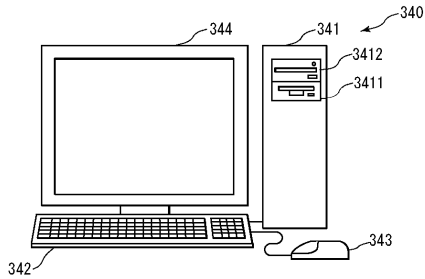
【図17】

用語 (辞書エントリ)	項目	言語				
		日本語	韓国語	中国語	英語	
大阪	表記	大阪	오사카	大阪	Osaka	--
	読み	おおさか	Osaka	Daban	ousa:ka:	--
	音響モデル		----	----	----	--
	アクセント	4E-70型	----	Da4ban3	óusaka	--
	他	----	----	<声調> 四声三声	----	--
東京	表記	東京	도쿄	東京	Tokyo	--
	読み	とうきょう	----	Tong1jing1	----	--
	音響モデル	----	----	----	----	--
	アクセント	----	----	----	tóukyó	--
	他	----	----	----	----	--
名古屋	表記	名古屋	----	----	----	--
	読み	----	----	----	----	--
	音響モデル	----	----	----	----	--
	他	----	----	----	----	--

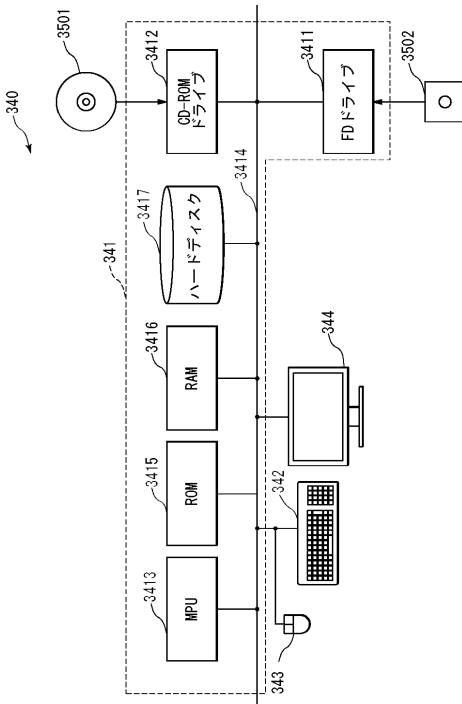
【図18】



【図19】



【図20】



---

フロントページの続き

(72)発明者 木村 法幸

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

(72)発明者 堀 智織

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

Fターム(参考) 5B091 AA03 CB12 CB32 CC15 CD03

5D015 GG00