IN THE UNITED STATES DISTRICT COURT
FOR THE EASTERN DISTRICT OF TEXAS
MARSHALL DIVISION

VB ASSETS, LLC, Plaintiff, v. SAMSUNG ELECTRONICS CO., LTD. and SAMSUNG ELECTRONICS AMERICA, INC, Defendants.	C.A. No. 2:24-cv-828 **JURY TRIAL DEMANDED**

NOTICE OF SUBPOENA TO AT&T COMMUNICATIONS, LLC.

PLEASE TAKE NOTICE THAT, pursuant to Rule 45(a)(4) of the Federal Rules of Civil Procedure, Defendants Samsung Electronics America, Inc. and Samsung Electronics Co., Ltd., by and through their undersigned counsel, will serve the attached document subpoena on AT&T Communications, LLC.

Dated: June 12, 2025

By: /s/ *Lora Krsulich*

Melissa R. Smith
Texas Bar No. 24001351
GILLIAM & SMITH, LLP
303 South Washington Avenue
Marshall, Texas 75670
Telephone: (903) 934-8450
Facsimile: (903) 934-9257
melissa@gilliamsmithlaw.com

Douglas J. Kline
Srikanth K. Reddy
Brian T. Drummond
GOODWIN PROCTER LLP
100 Northern Avenue
Boston, Massachusetts 02210
Telephone: (617) 570-1000
Facsimile: (617) 523-1231
dkline@goodwinlaw.com
sreddy@goodwinlaw.com
bdrummond@goodwinlaw.com

Darryl M. Woo
GOODWIN PROCTER LLP
525 Market Street, 32nd Floor
San Francisco, California 94105
Telephone: (415) 733-6000
Facsimile: (415) 677-9041
dwoo@goodwinlaw.com

Patrick J. McCarthy
Matthew Ginther
GOODWIN PROCTER LLP
1900 N Street, NW
Washington, DC 20036
Telephone: (202) 346-4000
Facsimile: (202) 346-4444
pmccarthy@goodwinlaw.com
mginther@goodwinlaw.com

Naomi Birbach
GOODWIN PROCTER LLP
The New York Times Building
620 Eighth Avenue
New York, New York 10018
Telephone: (212) 813-8800

Facsimile: (212) 355-3333
nbirbach@goodwinlaw.com

Jesse Cheng
GOODWIN PROCTER LLP
601 Marshall Street
Redwood City, California 94063
Telephone: (650) 752-3100
Facsimile: (650) 853-1038
jessecheng@goodwinlaw.com

Lora Krsulich
GOODWIN PROCTER LLP
601 South Figueroa Street, Suite 4200
Los Angeles, California 90017
Telephone: (213) 426-2500
Facsimile: (213) 623-1673
lkrsulich@goodwinlaw.com

Attorneys for Defendants Samsung Electronics Co., Ltd. and Samsung Electronics America, Inc.

CERTIFICATE OF SERVICE

The undersigned hereby certifies that on June 12, 2025 a true and correct copy of the foregoing document was served on all counsel of record.

/s/ Lora J. Krsulich

UNITED STATES DISTRICT COURT

for the

Eastern District of Texas

VB ASSETS, LLC, *Plaintiff* v. SAMSUNG ELECTRONICS CO., LTD., and SAMSUNG ELECTRONICS AMERICA, INC. *Defendant*	))))))	Civil Action No. 2:24-cv-828

SUBPOENA TO PRODUCE DOCUMENTS, INFORMATION, OR OBJECTS OR TO PERMIT INSPECTION OF PREMISES IN A CIVIL ACTION

To: AT&T Communications c/o The Corporation Trust Company
Corporation Trust Center, 1209 Orange Street, Wilmington, DE 19801

(Name of person to whom this subpoena is directed)

☑ *Production:* **YOU ARE COMMANDED** to produce at the time, date, and place set forth below the following documents, electronically stored information, or objects, and to permit inspection, copying, testing, or sampling of the material: SEE ATTACHED SCHEDULE A

Place: Goodwin Procter 1900 N St NW Washington, D.C. 20036	Date and Time: TBD

☐ *Inspection of Premises:* **YOU ARE COMMANDED** to permit entry onto the designated premises, land, or other property possessed or controlled by you at the time, date, and location set forth below, so that the requesting party may inspect, measure, survey, photograph, test, or sample the property or any designated object or operation on it.

Place:	Date and Time:

The following provisions of Fed. R. Civ. P. 45 are attached – Rule 45(c), relating to the place of compliance; Rule 45(d), relating to your protection as a person subject to a subpoena; and Rule 45(e) and (g), relating to your duty to respond to this subpoena and the potential consequences of not doing so.

Date: June 12, 2025

CLERK OF COURT

OR

Signature of Clerk or Deputy Clerk | /s/ *Srikanth Reddy* — *Attorney's signature*

The name, address, e-mail address, and telephone number of the attorney representing *(name of party)* Samsung Elec. Co. and Samsung Electronics America, Inc., who issues or requests this subpoena, are:

Srikanth K. Reddy, 100 Northern Ave. Boston, MA 02210, sreddy@goodwinlaw.com, 617-570-1465

Notice to the person who issues or requests this subpoena

If this subpoena commands the production of documents, electronically stored information, or tangible things or the inspection of premises before trial, a notice and a copy of the subpoena must be served on each party in this case before it is served on the person to whom it is directed. Fed. R. Civ. P. 45(a)(4).

Civil Action No. 2:24-cv-828

PROOF OF SERVICE

(This section should not be filed with the court unless required by Fed. R. Civ. P. 45.)

I received this subpoena for *(name of individual and title, if any)* _____

on *(date)* _____ .

❒ I served the subpoena by delivering a copy to the named person as follows: _____

_____ on *(date)* _____ ; or

❒ I returned the subpoena unexecuted because: _____

_____ .

Unless the subpoena was issued on behalf of the United States, or one of its officers or agents, I have also tendered to the witness the fees for one day's attendance, and the mileage allowed by law, in the amount of $ _____ .

My fees are $ _____ for travel and $ _____ for services, for a total of $ 0.00 .

I declare under penalty of perjury that this information is true.

Date: _____

_____
Server's signature

_____
Printed name and title

_____
Server's address

Additional information regarding attempted service, etc.:

Federal Rule of Civil Procedure 45 (c), (d), (e), and (g) (Effective 12/1/13)

(c) Place of Compliance.

(1) ***For a Trial, Hearing, or Deposition.*** A subpoena may command a person to attend a trial, hearing, or deposition only as follows:
(A) within 100 miles of where the person resides, is employed, or regularly transacts business in person; or
(B) within the state where the person resides, is employed, or regularly transacts business in person, if the person
(i) is a party or a party's officer; or
(ii) is commanded to attend a trial and would not incur substantial expense.

(2) ***For Other Discovery.*** A subpoena may command:
(A) production of documents, electronically stored information, or tangible things at a place within 100 miles of where the person resides, is employed, or regularly transacts business in person; and
(B) inspection of premises at the premises to be inspected.

(d) Protecting a Person Subject to a Subpoena; Enforcement.

(1) ***Avoiding Undue Burden or Expense; Sanctions.*** A party or attorney responsible for issuing and serving a subpoena must take reasonable steps to avoid imposing undue burden or expense on a person subject to the subpoena. The court for the district where compliance is required must enforce this duty and impose an appropriate sanction—which may include lost earnings and reasonable attorney's fees—on a party or attorney who fails to comply.

(2) ***Command to Produce Materials or Permit Inspection.***
(A) *Appearance Not Required.* A person commanded to produce documents, electronically stored information, or tangible things, or to permit the inspection of premises, need not appear in person at the place of production or inspection unless also commanded to appear for a deposition, hearing, or trial.
(B) *Objections.* A person commanded to produce documents or tangible things or to permit inspection may serve on the party or attorney designated in the subpoena a written objection to inspecting, copying, testing, or sampling any or all of the materials or to inspecting the premises—or to producing electronically stored information in the form or forms requested. The objection must be served before the earlier of the time specified for compliance or 14 days after the subpoena is served. If an objection is made, the following rules apply:
(i) At any time, on notice to the commanded person, the serving party may move the court for the district where compliance is required for an order compelling production or inspection.
(ii) These acts may be required only as directed in the order, and the order must protect a person who is neither a party nor a party's officer from significant expense resulting from compliance.

(3) ***Quashing or Modifying a Subpoena.***
(A) *When Required.* On timely motion, the court for the district where compliance is required must quash or modify a subpoena that:
(i) fails to allow a reasonable time to comply;
(ii) requires a person to comply beyond the geographical limits specified in Rule 45(c);
(iii) requires disclosure of privileged or other protected matter, if no exception or waiver applies; or
(iv) subjects a person to undue burden.
(B) *When Permitted.* To protect a person subject to or affected by a subpoena, the court for the district where compliance is required may, on motion, quash or modify the subpoena if it requires:
(i) disclosing a trade secret or other confidential research, development, or commercial information; or
(ii) disclosing an unretained expert's opinion or information that does not describe specific occurrences in dispute and results from the expert's study that was not requested by a party.
(C) *Specifying Conditions as an Alternative.* In the circumstances described in Rule 45(d)(3)(B), the court may, instead of quashing or modifying a subpoena, order appearance or production under specified conditions if the serving party:
(i) shows a substantial need for the testimony or material that cannot be otherwise met without undue hardship; and
(ii) ensures that the subpoenaed person will be reasonably compensated.

(e) Duties in Responding to a Subpoena.

(1) ***Producing Documents or Electronically Stored Information.*** These procedures apply to producing documents or electronically stored information:
(A) *Documents.* A person responding to a subpoena to produce documents must produce them as they are kept in the ordinary course of business or must organize and label them to correspond to the categories in the demand.
(B) *Form for Producing Electronically Stored Information Not Specified.* If a subpoena does not specify a form for producing electronically stored information, the person responding must produce it in a form or forms in which it is ordinarily maintained or in a reasonably usable form or forms.
(C) *Electronically Stored Information Produced in Only One Form.* The person responding need not produce the same electronically stored information in more than one form.
(D) *Inaccessible Electronically Stored Information.* The person responding need not provide discovery of electronically stored information from sources that the person identifies as not reasonably accessible because of undue burden or cost. On motion to compel discovery or for a protective order, the person responding must show that the information is not reasonably accessible because of undue burden or cost. If that showing is made, the court may nonetheless order discovery from such sources if the requesting party shows good cause, considering the limitations of Rule 26(b)(2)(C). The court may specify conditions for the discovery.

(2) ***Claiming Privilege or Protection.***
(A) *Information Withheld.* A person withholding subpoenaed information under a claim that it is privileged or subject to protection as trial-preparation material must:
(i) expressly make the claim; and
(ii) describe the nature of the withheld documents, communications, or tangible things in a manner that, without revealing information itself privileged or protected, will enable the parties to assess the claim.
(B) *Information Produced.* If information produced in response to a subpoena is subject to a claim of privilege or of protection as trial-preparation material, the person making the claim may notify any party that received the information of the claim and the basis for it. After being notified, a party must promptly return, sequester, or destroy the specified information and any copies it has; must not use or disclose the information until the claim is resolved; must take reasonable steps to retrieve the information if the party disclosed it before being notified; and may promptly present the information under seal to the court for the district where compliance is required for a determination of the claim. The person who produced the information must preserve the information until the claim is resolved.

(g) Contempt.
The court for the district where compliance is required—and also, after a motion is transferred, the issuing court—may hold in contempt a person who, having been served, fails without adequate excuse to obey the subpoena or an order related to it.

For access to subpoena materials, see Fed. R. Civ. P. 45(a) Committee Note (2013).

SCHEDULE A

DEFINITIONS

The following definitions shall apply to each request for production unless otherwise explicitly stated:

1. The term "VB Assets" and "VoiceBox" means VB Assets, LLC, and VoiceBox Technologies, and includes any past and present parents, predecessors, successors, subsidiaries, affiliates, divisions, associated organizations, joint ventures, and all present and former officers, directors, trustees, employees, staff members, agents, or other representatives, including counsel and patents agents, in any country.

2. The terms "you", "your," and "AT&T" mean "AT&T Communications, LLC" and "AT&T Corporation."

3. "Defendants" and "Samsung" mean Defendants Samsung Electronics America, Inc., and Samsung Electronics Co., Ltd., collectively.

4. The term "Amazon Litigation" refers to the action, *VB Assets, LLC v. Amazon, Inc. et al.*, 1:19-cv-01410-MN (D. Del.).

5. The term "Patents-In-Suit" shall refer, collectively, to U.S. Patent Nos. 8,073,681 ("the '681 patent"); 8,515,765 ("the '765 patent"); 10,510,341 ("the '341 patent"); 7,818,176 ("the '176 patent"); 8,886,536 ("the '536 patent"); and 10,755,699 ("the '699 patent).

6. The term "prior art" means, as of the filing date of the Patents-In-Suit, any article, poster, abstract, chapter, display, slides, or other printed publication that discloses, or a use, sale, or offer for sale of a system or device disclosed or claimed in the Patents-In-Suit or that practices or could be used to practice, the alleged inventions or portions of the alleged inventions disclosed or claimed in the Patents-In-Suit or any other thing or activity which could be or could have been

relied on by the United States Patent Office or a Court for an anticipation or obviousness determination of the Patents-In-Suit, including any and all patents, patent applications and/or publications prepared before the filing date of the Patents-In-Suit.

7. The term "related patents/applications" means any and all patents, patent applications and/or patent publications in the same patent family or families as the Patents-In-Suit and foreign counterparts; by way of example only, related patents/applications include any patent document that (i) claims priority from any of the Patents-In-Suit, (ii) is identified as priority for any of the Patents-In-Suit, (iii) claims priority to any application to which any of the Patents-In-Suit claims priority, (iv) shares the same or substantially similar specification, or a portion thereof, as any of the Patents-in-Suit (including any provisional, divisional, continuation, and continuation in-part applications), (v) is a reissue or reexamination of any of the Patents-in-Suit or any of the patents described in (i)-(iv), or is a foreign counterpart patent application or foreign counterpart patent to any of the patents or applications described in (i)-(iv).

8. The term "named inventor" shall mean one or more of the inventors named on any of the Patents-In-Suit.

9. The term "STT System" or Speech to Text System, means technology described in U.S. Patent No. 6,173,250, which is attached **Exhibit A,** and any related projects, services, or products.

10. The term "ASR," or Automated Speech Recognition, means technology enabling the use of computer hardware and software-based technique(s) to identify and process human voice, using systems and processes that interpret and convert received sound waves or other forms of spoken input to text and/or commands.

11. The term "NLU," or Natural Language Understanding, means technology enabling

the processing and comprehension of natural language, including speech and text, by software including, but not limited to, the process of mapping a sequence of words to an action to be performed by a computer.

12. The term "ASR/NLU technology" means the ASR and/or NLU technology of Speech-text transmit system and any products incorporating, applying, using, or relying on that technology that were designed, developed, made, used, promoted, demonstrated, beta tested, disclosed, publicly available, offered for sale, or sold before February 6, 2007, and any related projects, services, or products.

13. The term "Technology Related to Asserted Patents" means, within STT System: ASR/NLU technology, ASR/NLU hypothesis scoring, recognizing phonemes, using phonemes to recognize syllables, using syllables to recognize words, advertising applications in speech systems, promotional offers, short- and long-term contexts for processing utterances, NLU agents for different domains, ambiguity resolution, misrecognition correction and reinterpreting utterances based on subsequent utterances, tailoring system responses based on user conversation, and voice shopping.

14. The term "document" has the broadest meaning accorded that term by Fed. R. Civ. P. 34(a) and includes, but is not limited to, all of the items defined in Fed. R. Evid. 1001, and all preliminary and final drafts of any such item.

15. The term "all documents" means any and all documents that you can locate through a diligent search of all locations likely to contain documents requested herein and through reasonable inquiry of all persons likely to know of the existence of documents requested herein. A draft or non-identical copy is a separate document within the meaning of this term. Any comment or notation appearing on any document, and not a part of the original text, is to be considered a

separate "document."

16. The term "communication" shall refer to all written, oral, telephonic, or other inquiries, dialogues, discussions, conversations, interviews, correspondences, consultations, negotiations, agreements, understandings, meetings, letters, notes, telegrams, advertisements, press releases, computer mail, email and all other documents evidencing any verbal or nonverbal interaction between persons and/parties.

17. The term "person" is defined as any natural person or any legal entity, including, without limitation, any business or governmental entity or association.

18. The terms "relate to," "relates to," "related to," "relating to," "referring to," "pertaining to," "pertain to," and "regarding" mean constitute, include, comprise, consist of, refer, reflect, discuss, show, state, explain, contradict, provide context to, evidence, concern or be in any way logically or factually connected with the matter discussed or identified.

19. The terms "or" and "and" shall be read in the conjunctive and in the disjunctive wherever they appear, and neither of these words shall be interpreted to limit the scope of these requests.

20. The term "any" and "each" should be understood to include and encompass "all."

21. All pronouns shall be construed to refer to the masculine, feminine, or neutral gender, in singular or plural, as in each case makes the request more inclusive.

22. The use of a verb in any tense shall be construed as including the use of the verb in all other tenses.

23. The singular form of any word shall be deemed to include the plural. The plural form of any word shall be deemed to include the singular.

INSTRUCTIONS

1. The subpoena served upon you commands you to produce all responsive documents in your possession, custody or control, including your attorneys, agents, representatives, or employees.

2. Pursuant to Federal Rule of Civil Procedure 45(d)(1), you are to provide documents responsive to this subpoena as they are kept in the ordinary course of business, or can organize and label the documents to correspond with the categories set forth below.

3. Electronic records and computerized information must be produced in an intelligible format, together with a description of the system from which they were derived sufficient to permit rendering the records and information intelligible.

4. Selection of documents from the files and other sources and the numbering of such documents shall be performed in such a manner as to ensure that the source of each document may be determined, if necessary.

5. File folders with tabs or labels or directories of files identifying documents called for by these requests must be produced intact with such documents.

6. Documents attached to each other shall not be separated.

7. Should you seek to withhold any document based on some limitation of discovery (including, but not limited to, a claim of privilege), supply a list of the documents for which limitation of discovery is claimed, indicating:

   a. The identity of each document's author, writer, sender, or initiator;

   b. The identity of each document's recipient, addressee, or person for who it was intended;

   c. The date of creation or transmittal indicated on each document, or an estimate of

that date, indicated as such, if no date appears on the document;

d. The general subject matter as described on each document, or, if no such description appears, then some other description sufficient to identify the document; and

e. The claimed grounds for limitation of discovery (e.g., "attorney-client privilege").

8. If your response to a particular demand is a statement that you lack the ability to comply with that demand, you must specify whether the inability to comply is because the particular item or category never existed, has been destroyed, has been lost, misplaced, or stolen, or has never been, or is no longer, in your possession, custody, or control, in which case the name and address of any person or entity known or believed by you to have possession, custody, or control of that document or category of document must be identified.

9. To the extent permitted and authorized by law, these document requests shall be deemed continuing and require supplemental responses and production if you obtain additional documents between the time of initial production and the time of hearing or trial.

DOCUMENT REQUESTS

1. All documents and things produced by AT&T pursuant to a subpoena from Amazon in the Amazon Litigation.

2. All documents and things related to the STT System, including Technology Related to Asserted Patents, that were made, used, promoted, disclosed, demonstrated, publicly available, licensed (whether for free or for a fee) offered for sale, or sold before February 6, 2007, including but not limited to, products, devices, components, guides, schematics, specifications, software, hardware, firmware, applications, source code, prototypes, physical devices, test plans, test results, demonstrations, videos, white papers, memoranda, theses, publications, presentations, meeting minutes, manuals, business plans, and marketing plans.

3. All documents and things related to actual, planned, or hypothetical uses or applications of the STT System, including Technology Related to Asserted Patents, in the speech based interaction space.

4. All documents and things related to any beta testing of the STT System, including Technology Related to Asserted Patents, before February 6, 2007.

5. All documents identifying people most knowledgeable (whether current or former employees of AT&T) about the design, development, functionality, testing, operation, promotion, disclosure, licensing (whether for free or for a fee), demonstration, public use, sales offers, or sales of the STT System, including Technology Related to Asserted Patents, before February 6, 2007.

6. All documents related to sales or public demonstrations of the STT System, including Technology Related to Asserted Patents, before February 6, 2007 including, but not limited to, invoices, disclosures, and demonstration materials.

7. All documents identifying the third-party users of the STT System, including Technology Related to Asserted Patents, before February 6, 2007, including dates of use and extent of use. For clarity, "third-party users" as used in this topic includes any third parties that licensed (whether for free or for a fee), purchased, tested, modified, implemented, or otherwise used the STT System before February 6, 2007.

8. All documents, communications and things related to the asserted patents, related patent/applications, VB Assets, VoiceBox, or any named inventor relating to the subject matter of the asserted patents or VB Assets' business or to the STT System, including Technology Related to Asserted Patents.

EXHIBIT A

US006173250B1

(12) **United States Patent**
Jong

(10) **Patent No.: US 6,173,250 B1**
(45) **Date of Patent: Jan. 9, 2001**

(54) **APPARATUS AND METHOD FOR SPEECH-TEXT-TRANSMIT COMMUNICATION OVER DATA NETWORKS**

(75) Inventor: **Kenneth Jong**, Bridgewater, NJ (US)

(73) Assignee: **AT&T Corporation**, New York, NY (US)

(*) Notice: Under 35 U.S.C. 154(b), the term of this patent shall be extended for 0 days.

(21) Appl. No.: **09/089,855**

(22) Filed: **Jun. 3, 1998**

(51) **Int. Cl.**[7] **G10L 13/00**; G06F 17/28

(52) **U.S. Cl.** **704/3**; 704/269; 379/88.06; 379/88.16

(58) **Field of Search** 704/201, 221, 704/235, 260, 270, 500, 501, 249, 258, 272, 2, 3, 4; 379/88.05, 88.06, 88.16, 88.18

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,641,264		2/1987	Nitta et al.	704/4
4,706,212		11/1987	Toma	704/2
4,980,829		12/1990	Okajima et al.	704/5
5,029,200	*	7/1991	Haas et al.	379/88.18
5,293,584		3/1994	Brown et al.	704/277
5,384,701		1/1995	Stentiford et al.	704/3
5,426,583		6/1995	Uribe-Echebarria Diaz	704/2
5,477,450		12/1995	Takeda et al.	704/2
5,497,319		3/1996	Chong et al.	704/2
5,510,981		4/1996	Berger et al.	704/2
5,640,575		6/1997	Maruyama et al.	704/4
5,675,815		10/1997	Yamauchi et al.	707/530
5,677,835		10/1997	Carbonell et al.	704/8
5,687,383		11/1997	Nakayama et al.	704/2
5,696,879	*	12/1997	Cline et al.	704/260
5,696,980		12/1997	Brew	704/273
5,712,901		1/1998	Meermans	379/88.14
5,724,410	*	3/1998	Parvulescu et al.	704/258
5,857,170	*	1/1999	Kondo	704/266
5,875,422	*	2/1999	Eslambolchi et al.	704/3
5,884,262	*	3/1999	Wise et al.	704/270
5,911,129	*	6/1999	Towell	704/272
5,915,237	*	6/1999	Boss et al.	704/258
5,915,238	*	6/1999	Tjaden	704/260
5,933,805	*	8/1999	Boss et al.	704/249
5,940,796	*	8/1999	Matsumoto	704/260
5,949,854	*	9/1999	Sato	379/88.16
6,014,428	*	1/2000	Wolf	379/88.11
6,073,094	*	6/2000	Chang et al.	704/223

OTHER PUBLICATIONS

Magedi et al., "Proposals of MIDI coding and its application for audio authoring," IEEE International Conference on Multimedia Computing and Systems, pp. 305–314, Jul. 1998.*

* cited by examiner

Primary Examiner—David R. Hudspeth
Assistant Examiner—Martin Lerner
(74) *Attorney, Agent, or Firm*—Oliff & Berridge, PLC

(57) **ABSTRACT**

An apparatus and method for speech-text-transmit communication over data networks includes speech recognition devices and text to speech conversion devices that translate speech signals input to the terminal into text and text data received from a data network into speech output signals. The speech input signals are translated into text based on phonemes obtained from a spectral analysis of the speech input signals. The text data is transmitted to a receiving party over the data network as a plurality of text data packets such that a continuous stream of text data is obtained. The receiving party's terminal receives the text data and may immediately display the text data and/or translate it into speech output signals using the text to speech conversion device. The text to speech conversion device uses speech pattern data stored in a speech pattern database for synthesizing a human voice for playing of the speech output signals using a speech output device.

20 Claims, 7 Drawing Sheets

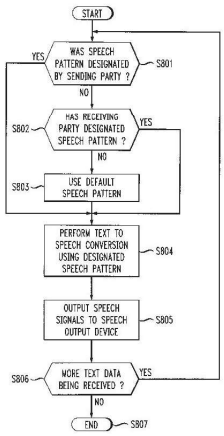


FIG. 1

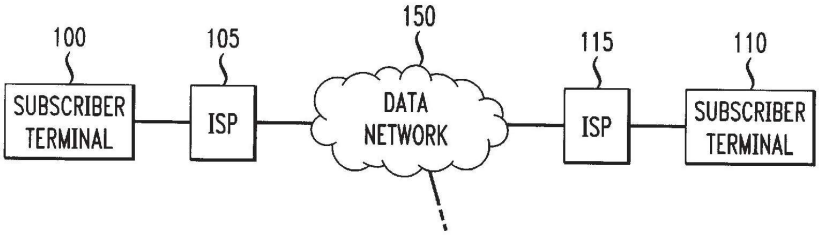

100
105
150
115
110
SUBSCRIBER TERMINAL
ISP
DATA NETWORK
ISP
SUBSCRIBER TERMINAL

FIG. 2

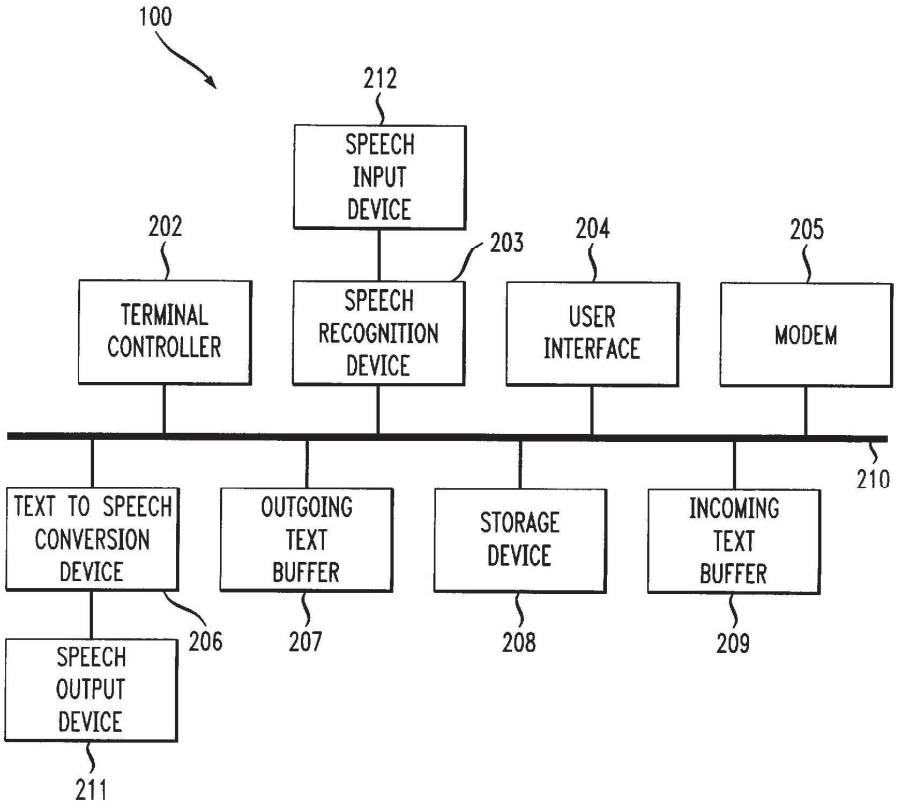

100
212
SPEECH INPUT DEVICE
202
203
204
205
TERMINAL CONTROLLER
SPEECH RECOGNITION DEVICE
USER INTERFACE
MODEM
210
TEXT TO SPEECH CONVERSION DEVICE
OUTGOING TEXT BUFFER
STORAGE DEVICE
INCOMING TEXT BUFFER
206
207
208
209
SPEECH OUTPUT DEVICE
211

FIG. 3

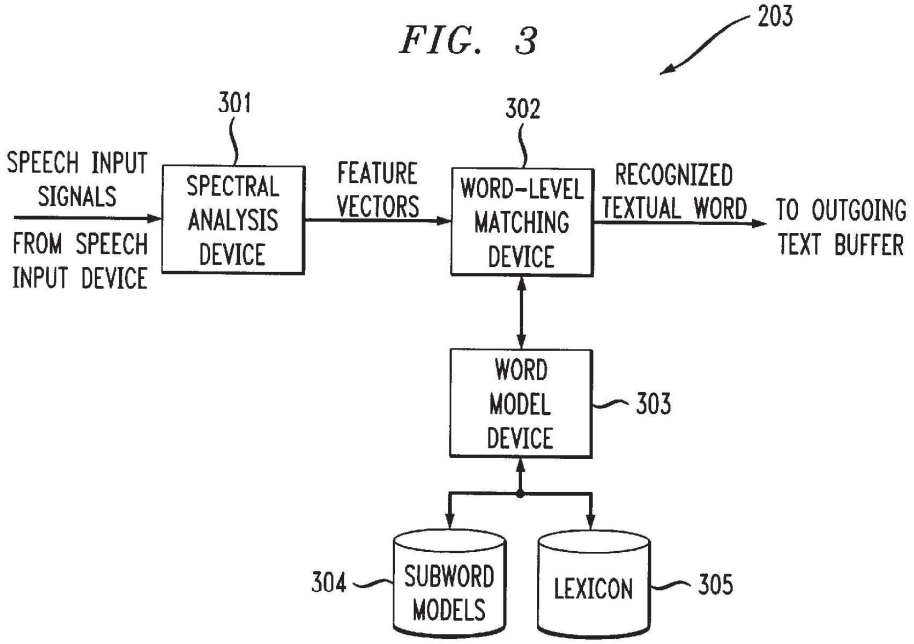

203
301
302
SPEECH INPUT SIGNALS
FROM SPEECH INPUT DEVICE
SPECTRAL ANALYSIS DEVICE
FEATURE VECTORS
WORD-LEVEL MATCHING DEVICE
RECOGNIZED TEXTUAL WORD
TO OUTGOING TEXT BUFFER
WORD MODEL DEVICE
303
304
SUBWORD MODELS
LEXICON
305

FIG. 4

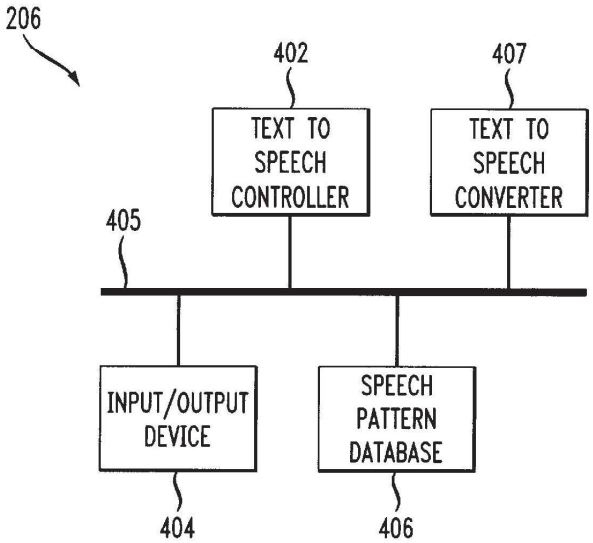

206
402
407
TEXT TO SPEECH CONTROLLER
TEXT TO SPEECH CONVERTER
405
INPUT/OUTPUT DEVICE
SPEECH PATTERN DATABASE
404
406

FIG. 5

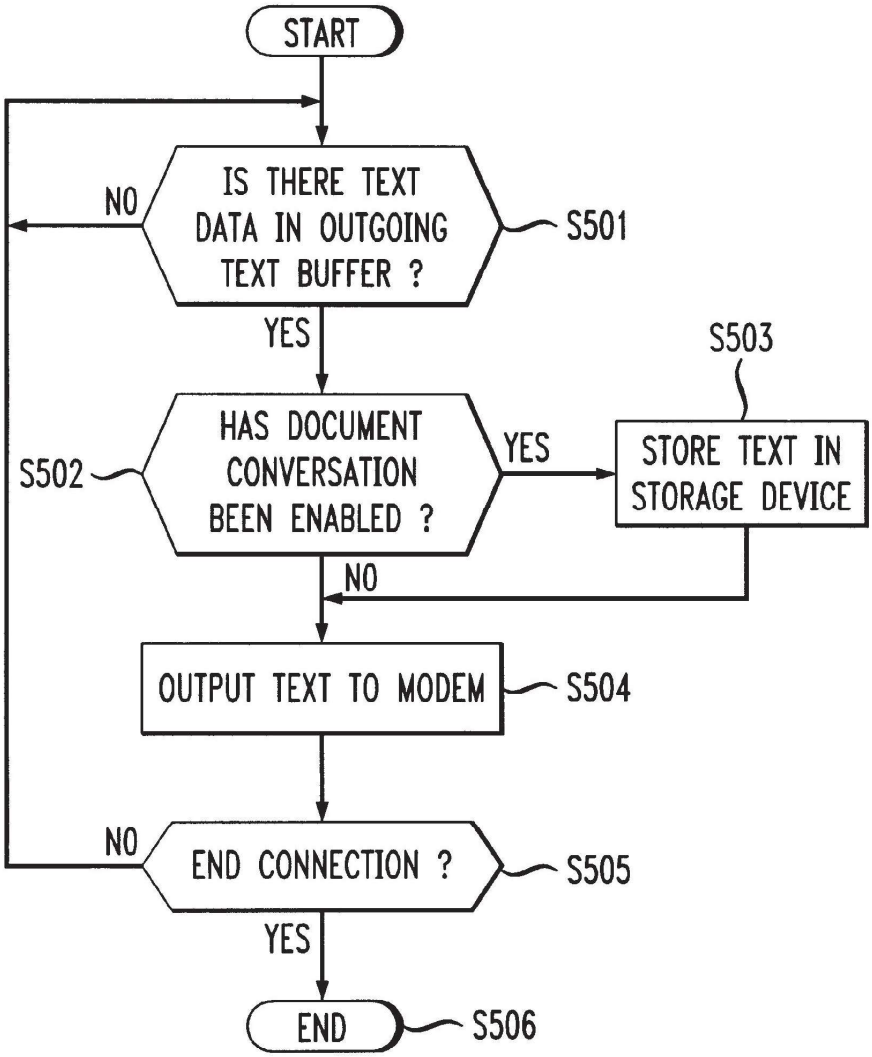

START
IS THERE TEXT DATA IN OUTGOING TEXT BUFFER ?
S501
NO
YES
HAS DOCUMENT CONVERSATION BEEN ENABLED ?
S502
YES
NO
S503
STORE TEXT IN STORAGE DEVICE
OUTPUT TEXT TO MODEM
S504
END CONNECTION ?
S505
NO
YES
END
S506

FIG. 6

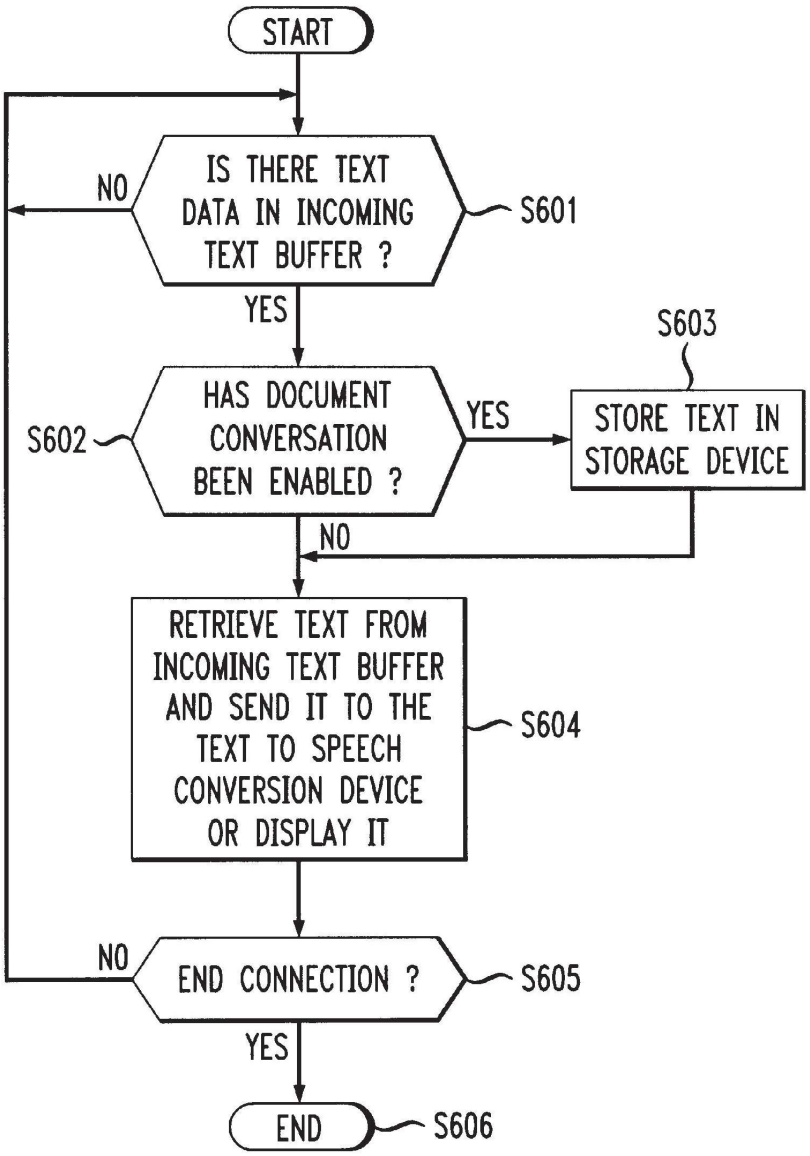

START
IS THERE TEXT DATA IN INCOMING TEXT BUFFER ?
S601
NO
YES
HAS DOCUMENT CONVERSATION BEEN ENABLED ?
S602
YES
NO
S603
STORE TEXT IN STORAGE DEVICE
RETRIEVE TEXT FROM INCOMING TEXT BUFFER AND SEND IT TO THE TEXT TO SPEECH CONVERSION DEVICE OR DISPLAY IT
S604
END CONNECTION ?
S605
NO
YES
END
S606

FIG. 7

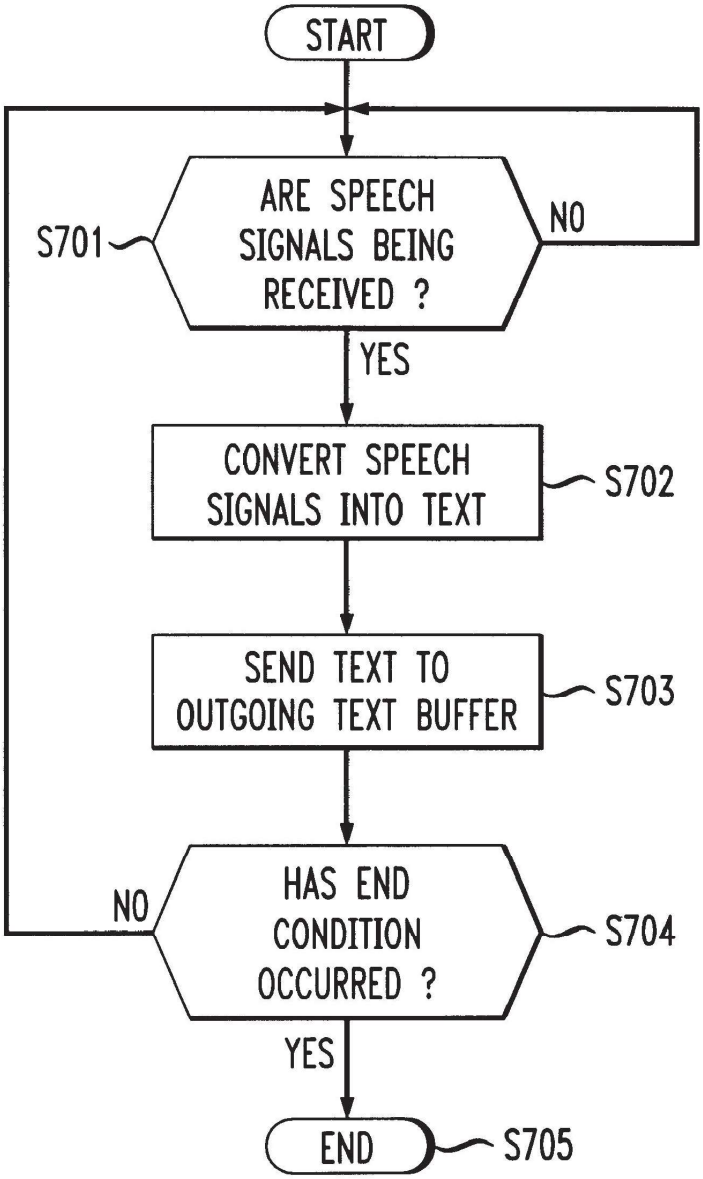
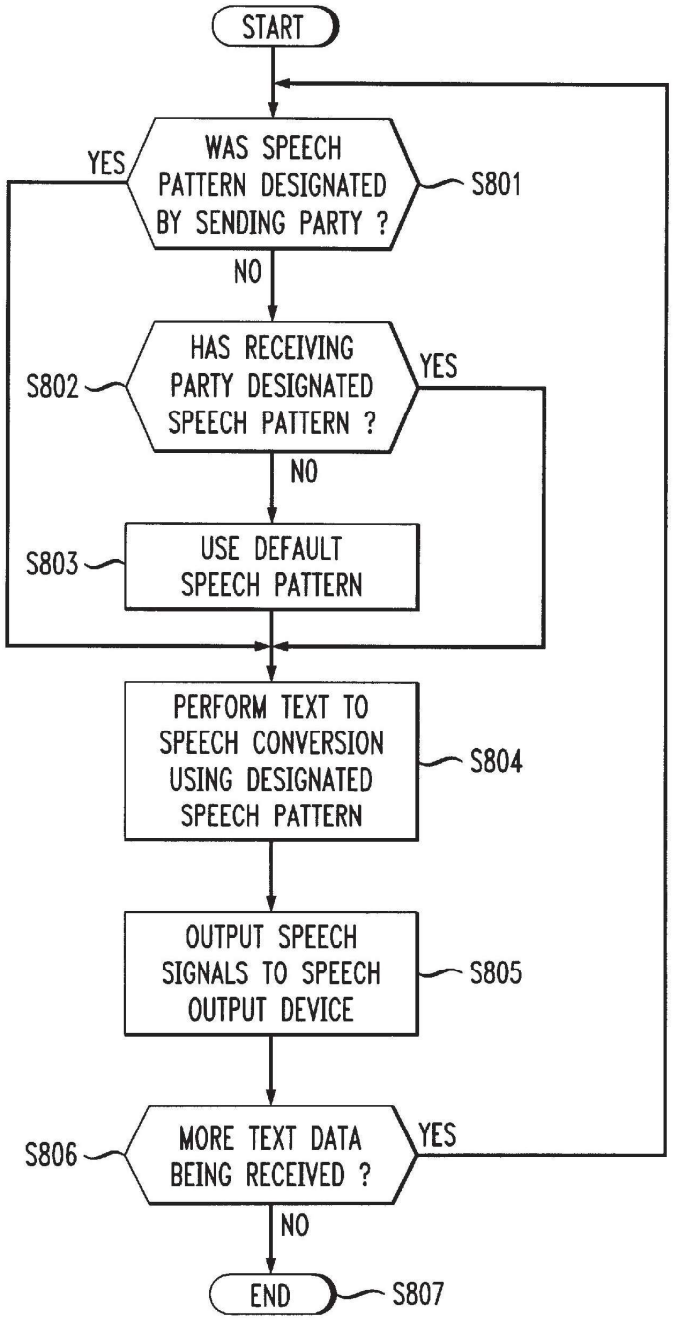

START
ARE SPEECH SIGNALS BEING RECEIVED ?
S701
NO
YES
CONVERT SPEECH SIGNALS INTO TEXT
S702
SEND TEXT TO OUTGOING TEXT BUFFER
S703
HAS END CONDITION OCCURRED ?
S704
NO
YES
END
S705

FIG. 8


START
WAS SPEECH PATTERN DESIGNATED BY SENDING PARTY ?
S801
YES
NO
HAS RECEIVING PARTY DESIGNATED SPEECH PATTERN ?
S802
YES
NO
USE DEFAULT SPEECH PATTERN
S803
PERFORM TEXT TO SPEECH CONVERSION USING DESIGNATED SPEECH PATTERN
S804
OUTPUT SPEECH SIGNALS TO SPEECH OUTPUT DEVICE
S805
MORE TEXT DATA BEING RECEIVED ?
S806
YES
NO
END
S807

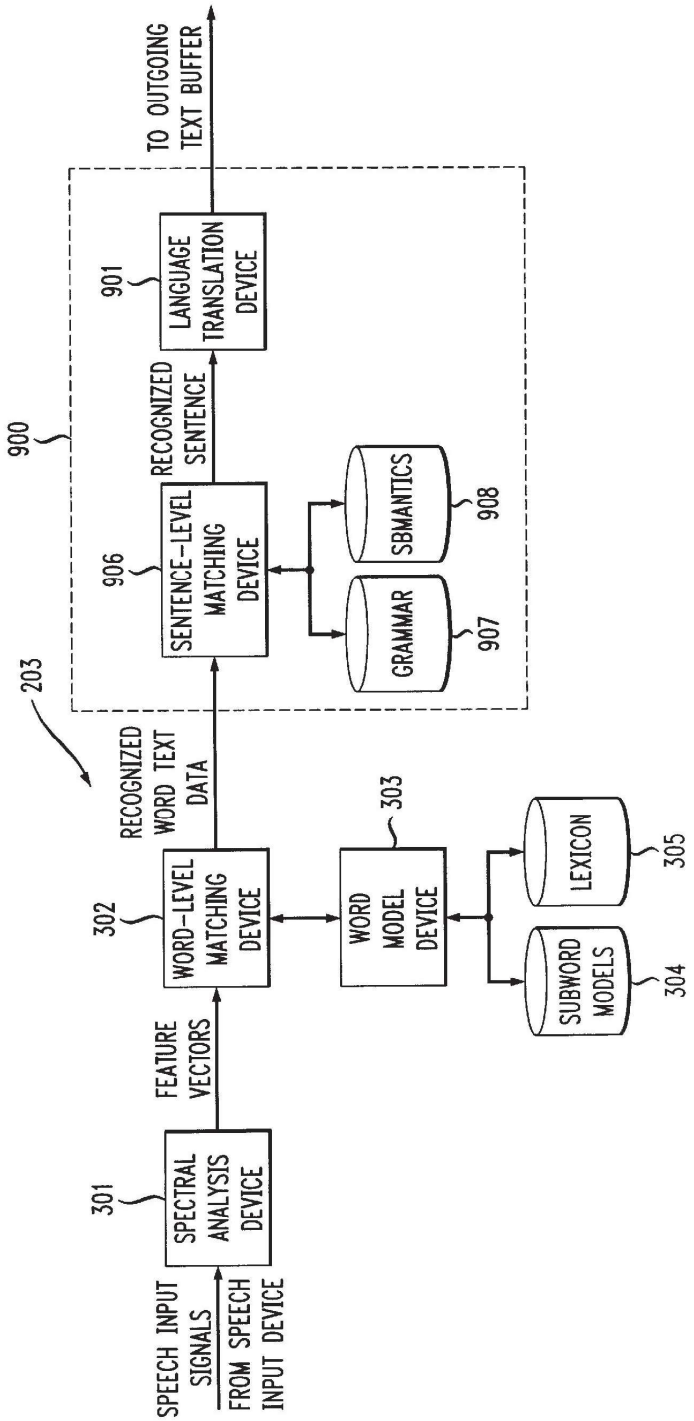

SPEECH INPUT SIGNALS
FROM SPEECH INPUT DEVICE
301
SPECTRAL ANALYSIS DEVICE
FEATURE VECTORS
302
WORD-LEVEL MATCHING DEVICE
303
WORD MODEL DEVICE
SUBWORD MODELS
304
LEXICON
305
203
RECOGNIZED WORD TEXT DATA
900
906
SENTENCE-LEVEL MATCHING DEVICE
GRAMMAR
907
SBMANTICS
908
RECOGNIZED SENTENCE
901
LANGUAGE TRANSLATION DEVICE
TO OUTGOING TEXT BUFFER

FIG. 9

APPARATUS AND METHOD FOR SPEECH-TEXT-TRANSMIT COMMUNICATION OVER DATA NETWORKS

BACKGROUND OF THE INVENTION

1. Field of Invention

The invention pertains to an apparatus and method for providing real time communication over a data network.

2. Description of Related Art

With the ever increasing benefits of computerization to every day life, more people are logging-on to the Internet to obtain information and to communicate with other people around the world. This communication typically takes the form of textual messages sent either by electronic mail, text messages sent to other users currently logged-on to designated "chat rooms" or sent through Internet messaging. These text messages must be manually input via a computer keyboard and sent to the other users as blocks of text. This causes a great delay in the transmission and reception of messages since the speed at which communication takes place is dependent on the users' typing ability.

Recently, voice telephony systems have been devised which use voice codes (coder/decoders) to achieve actual voice communication over data networks. These systems take voice input signals, digitize these signals, and encode them for transmission as digital data. This data is transmitted to a receiving station where the coded data is decoded and output to the receiving party. These systems are straddled with significant sound degradation (delays and distortions) due to the asynchronous nature of the data network and the continuous streaming of the data packets necessary for the transmission of actual voice data. The timing of the data packets used in such voice telephony systems requires large bandwidth and cannot tolerate undesirable delays. Therefore, the use of standard modem lines having small bandwidths in voice telephony systems results in degradation of the reconstructed sound.

Furthermore, these voice telephony systems do not allow for anonymity of the parties since the party's actual voice is transmitted. Such anonymity is generally valued by users of the Internet because it provides a sense of privacy and security. In addition, automatic documentation of conversations are not possible and the voice telephony is limited to the particular languages spoken by the parties. Thus, new technology is needed to achieve acceptable voice communication over data networks.

SUMMARY OF THE INVENTION

The present invention is directed to an apparatus and method for providing speech-text-transmit communication over data networks. With the present invention, speech input signals are converted into text data through the use of speech recognition and transmitted over a data network to a designated receiving party. The text data may be converted back into speech output signals at the receiving party's terminal. The speech output signals are obtained using a synthesized speech pattern.

Since text data packets representing speech are streaming at a lower data rate when compared with voice telephony data packets, the transmission of the text data packets may

be performed at a lower speed and therefore a lower bandwidth. Thus, a realtime continuous text stream and synthesized speech output may be realized using a lower data rate in data networks. In addition, the text data may be displayed to the receiving party or stored in a storage device for later review and reproduction. In this way, the communication may be documented.

Further, the speech pattern data of synthesized speech used to output the converted text data may be designated by the sending and/or receiving parties. In this way, the system is tailorable to the particular preferences of the parties involved in the communication.

Further, the present invention provides an apparatus for translating the text data received in one language into text data of another language. In this way, communication may be facilitated in the parties' own languages thereby reducing the chances of misinterpretation.

BRIEF DESCRIPTION OF THE DRAWINGS

The preferred embodiments of this invention are described in detail with reference to the following figures, wherein like numerals represent like elements, and wherein:

FIG. **1** is a block diagram of a speech-text-transmit system;

FIG. **2** is a block diagram of the subscriber terminal of FIG. **1**;

FIG. **3** is a block diagram of the speech recognition device of FIG. **2**;

FIG. **4** is a block diagram of the text to speech conversion device of FIG. **3**;

FIG. **5** is a flowchart of the operation of the terminal controller of FIG. **2** when monitoring the output text buffer for text data to be output via the modem;

FIG. **6** is a flowchart of the operation of the terminal controller of FIG. **2** when monitoring the input text buffer for text data received through the modem;

FIG. **7** is a flowchart of the operation of the speech recognition device of FIG. **3**;

FIG. **8** is a flowchart of the operation of the text to speech conversion device of FIG. **2**; and

FIG. **9** is a block diagram of the speech recognition device with an additional language translation device.

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

FIG. **1** is a block diagram of an exemplary embodiment of a speech-text-transmit (STT) system. The STT system includes a subscriber terminal **100** and a subscriber terminal **110** in connection with a data network **150** through Internet Service Providers (ISPs) **105** and **115**, for example.

The data network **150** may be any type of data network used for communication and data transfer. Examples include the Internet, local area networks (LANs), wide area networks (WANs), and the like. For purposes of explanation of the present invention, the data network **150** is considered to be the Internet.

The subscriber terminals **100** and **110** may be personal computers connected to the data network **150** via modem connections, computer terminals connected via a LAN or

WAN, WebTV™ devices, telephones (wired or wireless) connected to computing gateways with access to the Internet, and the like. For purposes of explanation of the present invention, subscriber terminals **100** and **110** are considered to be personal computers connected to the data network via modem connections.

With the present invention, when subscriber terminal **100** wishes to communicate with subscriber terminal **110**, assuming both subscriber terminals are logged onto the data network **150** through their respective ISPs **105** and **115**, the subscriber terminal **100** designates subscriber terminal **110** as the recipient of further communications. The subscriber terminal **100** may then initiate communication with subscriber terminal **110** via the data network **150** by receiving audio input signals from the user and converting them into textual representations of the audio input signals. These textual representations are then sent to the subscriber terminal **110** via the data network **150**. Subscriber terminal **110** may then respond in kind. In this way, realtime communication is achieved by the sending of continuous streams of text data over the data network **150** to and from the subscriber terminals.

FIG. **2** is a block diagram of subscriber terminal **100**. The subscriber terminal **110** is similarly equipped. As shown in FIG. 2, the subscriber terminal **100** includes a terminal controller **202**, a speech recognition device **203**, a user interface **204**, a modem **205**, a text to speech conversion device **206**, an outgoing text buffer **207**, a storage device **208** and an incoming text buffer **209**. These devices are connected to one another by way of bus **210**. The terminal controller **202** controls the flow of information along the bus **210** to each of the devices.

A speech input device **212** is connected to the terminal **100** through the speech recognition device **203**. The speech input device **212** includes, for example, a microphone, telephone handset, telephone headset, and the like. The speech input device **212** detects a user's speech and converts it into speech signals that are sent to the speech recognition device **203**.

A speech output device **211** is connected to terminal **100** through the text to speech conversion device **206**. The speech output device **211** includes, for example, speakers, amplifiers, and the like for providing audible output signals. Speech signals sent from the text to speech conversion device **206** are output by the speech output device **211**.

The user interface **204** includes, for example, a keyboard, a display, a pointing device, and the like. The user interface **204** handles user requests by receiving, for example, keyboard entries combined with the actions of a pointing device, such as a computer mouse or trackball, to establish and modify the communications process.

The modem **205** includes any type network interface that facilitates the transmission and reception of data from the terminal **100** to the data network **150** and vice versa. This includes both external and internal modem devices of various baud rates. The modem **205**, in a preferred embodiment, is intended to be of the generally available type that can be purchased in most retail electronics stores. However, specialized modems are also intended to be within the spirit and scope of this invention. In alternative embodiments, the modem may be replaced by, for example, a local area network card, and the like, that allows the terminal to interact as part of a computer network. Any such device that allows for the communication between two or more terminals is intended to be represented by the modem **205**.

The storage device **208** may be any type of storage device that provides for the storage of electronic data. For example, the storage device **208** may be a hard disk, floppy disk, magnetic tape, writeable CD, zip disk, and the like. The storage device **208** provides storage for data and, in particular, provides storage for documenting communications using the present invention.

The outgoing and incoming text buffers **207** and **209** may be any type of memory device used for the temporary storage of information. The text buffers **207** and **209** may be a RAM, hard disk, floppy disk, magnetic tape, cache and the like. The text buffers **207** and **209** serve to provide a temporary storage of text data received and to be transmitted over the data network **150** via modem **205**. In this way, a continuous stream of text data may be received from and transmitted to the receiving party (either subscriber terminal **100** or **110**).

The terminal controller **202** provides the controlling activities to set up communications protocols which, for example, include Internet Protocol (IP) addresses of communications parties and the Universal Resource Locator (URL) address of the hosting site, to initiate the communications process. When transmitting text data, the terminal controller **202** coordinates the timing sequence and signal flow from the speech input at the speech input device **212**, through the speech recognition device **203**, the outgoing text buffer **207**, the storage device **208** if needed, and the modem **205**, to output a text stream to the Internet. When receiving text data, the terminal controller **202** coordinates the timing sequence and signal flow from the modem **205**, through the incoming text buffer **209**, the text to speech conversion device **206** and the speech output device **211**.

The subscriber terminals **100** and **110** are connected to the data network **150** via their modems **205** through Internet Service Providers (ISPs) **105** and **115**, respectively. The subscriber terminals **100** and **110** log onto the data network **150** by establishing modem connections to their respective ISPs **105** and **115** and enabling browsing software such as Netscape™ or Microsoft Internet Explorer™ (generally supplied by the particular ISP or obtainable through third parties). When logging onto the ISPs **105** and **115**, the subscriber terminals **100** and **110** are assigned Internet Protocol (IP) addresses within the data network **150**. Communication with other subscribers through the data network **150** is performed using these IP addresses and may include a host, as is well known in the art.

One type of communication generally used by the subscriber terminals **100** and **110** is chat. In this type of communication, the subscriber terminals **100** and **110** log onto their respective ISPs **105** and **115**, enable their browser software and enter designated chat sites by either locating them using search engines or entering the site's URL. In these chat sites, the subscriber terminals **100** and **110** are provided with a list of other chat participants currently logged onto that chat site. The operators of the subscriber terminals **100** and **110** may type messages to all of the other

chat participants or may designate certain ones of the chat participants with which to communicate.

In one embodiment, the subscriber terminals **100** and **110** are assumed to be logged onto a same chat site. In this chat site, the subscriber terminal **100** designates subscriber terminal **110** as the party with whom speech communication is desired. Once the subscriber terminal **110** is designated, the operator of the subscriber terminal **100** initiates communication with the operator of the subscriber terminal **110** by speaking into the speech input device **212**. The speech input is translated into speech signals by the speech input device **212**. The speech signals are then input to the speech recognition device **203** where they are converted into text data. This text data is then sent to the output text buffer **207** via bus **210**. The terminal controller **202** causes the outgoing text buffer **207** to be continuously emptied by transmitting the text data stored in the outgoing text buffer **207** to the IP address of the subscriber terminal **110** via the modem **205**, ISP **105**, data network **150** and ISP **115**.

The text data arriving at modem **205** of the subscriber terminal **110** is first stored temporarily in inpcoming text buffer **209**. From the incoming text buffer **209**, the text data can be displayed immediately or may be forwarded to the text to speech conversion device **206** of subscriber terminal **110** where the text data is converted into synthesized speech output signals using stored speech pattern data as will be described below. The operator of the subscriber terminal **110** may reply in the same manner, thereby enabling real time speech communication using text data transmission over the data network **150**.

FIG. **3** is a detailed block diagram of an exemplary speech recognition device **203** of FIG. **2**. As shown in FIG. **3**, the speech recognition device **203** includes a spectral analysis device **301**, a word-level matching device **302**, a word model device **303**, a subword models database **304**, and a lexicon database **305**.

When the speech input signals are received by the speech recognition device **203**, the spectral analysis device **301** receives the speech input signals and extracts feature vectors from them. The feature vectors are input to a word-level matching device **302** which compares the feature vectors against the word models retrieved by the word model device **303** to identify the words that make up the speech input signals.

The word model device **303** retrieves information from the subword model and lexicon databases **304** and **305**. The word model device **303** includes a listing of phonemes (speech sounds) which are used to identify the words in the speech input signals. The subword model database **304** contains word syllables that are correlated with the phonemes of the word model device **303**. The lexicon database **305** stores a dictionary of recognizable words.

The word model device **303** identifies the phonemes in the speech input signals and extracts the corresponding syllables from the subword model database **304**. The syllables are then checked against the lexicon data stored in the lexicon database **305**. In this way, the syllables that make up the various words in the speech input signals are grouped into the recognizable words identified using the lexicon database **305**. The identified words are sent to the outgoing text buffer **207**. In this way a continuous stream of text data is sent to the outgoing text buffer **207** and, in turn, is output to the modem **205** and transmitted to the receiving party via the data network **150**.

FIG. **4** is a block diagram of the text to speech conversion device **206** of FIG. **2**. When text data is received through the modem **205**, it is temporarily stored in the incoming text buffer **209**. The terminal controller **202** then causes the stored text data to be routed to the text to speech conversion device **206**. The text to speech controller **402** receives the text data through the input/output device **404** and routes the text data to the text to speech converter **407**. The text to speech converter **407** converts the text data into speech output signals using synthesized speech pattern data stored in speech pattern database **406**. Such translation of text into audio signals using synthesized speech pattern data is generally known in the art and the description is therefore omitted. These speech output signals are then output to the receiving party through the speech output device **211**.

The speech pattern database **406** contains a plurality of speech patterns that may be used to provide a synthesized speech output of the text data. These speech patterns may have various characteristics representing different types of persons. For example, one speech pattern may represent an adult male and another may represent an adult female. Alternatively, a speech pattern of the party actually sending the text data may be stored in the speech pattern database **406** in order to obtain a synthesized speech output simulating the sending party's actual voice.

Another feature of the invention is the ability to document the communications being transmitted to and from the parties involved in the communication. To enable this feature, the operator of the subscriber terminal **100** or **110** inputs a command through the user interface **204** to enable documenting of communications. When this feature is enabled, all text data that is either received or sent is also stored in the storage device **208** as a text data file. Thus, when the terminal controller **202** determines that text data is being received, a copy of the text data is stored in the storage device **208**. Likewise, when the terminal controller determines that text data is to be transmitted from the outgoing text buffer **207**, a copy of the text data is also stored in the storage device **208**. Depending on the source of the text data, the terminal controller **202** may also cause identifying text to be added that identifies the source of the text data. For example, in a communication between Steve and John, the terminal controller **202** will add the name "Steve" before all text being transmitted from the outgoing text buffer **207** and the name "John" before all text data being received from the modem **205**. In this way, the operator, "Steve", may retain a transcript of the communication for later reference.

The received text data may be displayed immediately to the operator and not translated to speech output signals at all. In this way, one party may communicate in text only mode while the other party may communicate using speech.

Using the above described speech-text-transmit system, realtime continuous communication over the data network **150** is realized using existing modem lines. Since text data does not require as high a data rate as voice telephony data packets, the speed at which the data must be sent is reduced, thereby allowing for conventional modem lines to transmit

the data without degradation. Additionally, since the audio output is of a synthesized speech, anonymity may also be preserved.

FIGS. **5** to **8** are flowcharts of the process described above. This process is from the viewpoint of only one subscriber terminal **100** or **110**. However, the same operations are performed by both subscriber terminals **100** and **110**. Also, while the flowchart assumes, for simplicity, that either text data is being received or speech data is being input, it is recognized that text data may be received and speech data may be input at the same time. Thus, translation from text to speech and from speech to text may be performed concurrently.

FIG. **5** is a flowchart of the operation of the terminal controller **202** when transmitting text data. In step S**501**, the terminal controller **202** determines whether text data is to be transmitted by determining if there is text data present in the outgoing text buffer **207**. If text data is present in the outgoing text buffer **207**, the terminal controller **202** proceeds to step S**502**; otherwise the terminal controller **202** returns to step S**501**. In step S**502**, the terminal controller **202** determines whether the operator has indicated that the conversation should be documented. If the conversation is to be documented, the terminal controller **202** goes to step S**503**; otherwise the terminal controller **202** goes to step S**504**. In step S**503**, the terminal controller **202** stores a copy of the text data in the storage device **208** for later retrieval and goes to step S**504**.

In step S**504**, the terminal controller **202** outputs the text data from the outgoing text buffer **207** to the modem **205** and goes to step S**505**. The modem **205** transmits the text data to the receiving party **110** via the communications links to ISP **105**, data network **150** and ISP **115**. In step S**505**, the terminal controller **202** determines if an end condition has arisen. This end condition is, for example, the termination of the conversation and the like. If the end condition has occurred, the terminal controller **202** goes to step S**506** and ends; otherwise, the terminal controller **202** returns to step S**501**.

FIG. **6** is a flowchart of the operation of the terminal controller **202** when text data is being received via the modem **205** and stored in input text buffer **209**. In step S**601**, the terminal controller **202** determines if there is text data stored in the incoming text data buffer **209**. If there is text data stored in the incoming text buffer **209**, the terminal controller **202** goes to step S**602**; otherwise, the terminal controller **202** returns to step S**601**. In step S**602**, the terminal controller **202** determines whether the operator has designated that the conversation is to be documented. If the conversation is to be documented, the terminal controller **202** goes to step S**603**; otherwise, the terminal controller **202** goes to step S**604**. In step S**603**, the terminal controller **202** stores a copy of the text data received by the incoming text buffer **209** in the storage device **208** for later retrieval, and goes to step S**604**.

In step S**604**, the terminal controller **202** retrieves the text data from the input text buffer and either sends it to the text to speech conversion device **206** or simply displays the text, depending on the settings input by the operator. If the operator has chosen to only display the incoming text, the text data is displayed through the user interface **204**. Otherwise, the text data is sent to the text to speech conversion device **206** for conversion into synthesized speech output signals.

Next, in step S**605**, the terminal controller **202** determines whether an end condition has occurred. This end condition may be, for example, the termination of the conversation and the like. If an end condition has occurred, the terminal controller **202** goes to step S**606** and ends; otherwise, control returns to step S**601**.

FIG. **7** is a flowchart showing the operation of the speech recognition device **203**. In step S**701**, the speech recognition device **203** determines if speech signals are being received via the speech input device **212**. If speech signals are being received, control continues to step S**702**; otherwise, control returns to step S**701** until speech signals are detected.

In step S**702**, the speech recognition device **203** converts the speech signals that are being received into text data as described above. Control then continues to step S**703** where the speech recognition device **203** outputs the text data to the outgoing text buffer **207**. Next, in step S**704**, the speech recognition device **203** determines if an end condition has occurred. The end condition may be, for example, the termination of the conversation and the like. If an end condition has occurred, control goes to step S**705** and ends; otherwise, control returns to step S**701**.

FIG. **8** is a flowchart of the operation of the text to speech conversion device **206**. When text data is received and stored in the incoming text buffer **209**, the terminal controller **202** routes the text data to the text to speech conversion device **206**, if the text to speech function is enabled by the user. Then, in step S**801**, the text to speech controller **402** of the text to speech conversion device **206**, determines if the speech pattern to be used has been designated by the sending party. This may be determined by analyzing the signal to determine if a header signal identifying the speech pattern is present. If the sending party has designated the speech pattern, control jumps to step S**804**; otherwise, control continues to step S**802**.

In step S**802**, the text to speech controller **402** determines if the receiving party has designated the speech pattern to be used through the user interface **204**. If the receiving **10** party has designated the speech pattern, control jumps to step S**804**; otherwise, control continues to step S**803** where a default speech pattern is used.

In step S**804**, the text to speech controller **402** causes the text to speech conversion to be performed by the text to speech converter **407** using the designated speech pattern. Next, in step S**805**, the text to speech controller **402** outputs the output speech signals, generated by the text to speech converter **407** in step S**804**, to the speech output device **211**. Then, in step S**806**, the text to speech controller **402** determines if there is more text data that needs to be converted into speech. If there is, then control returns to step S**801**; otherwise, control goes to step S**807** and ends.

FIG. **9** shows a third embodiment in which the speech recognition device **203** outputs text data in a selected language. In this embodiment, the speech recognition device **203** further receives commands through the user interface **204** indicating the selected language. The speech recognition device **203** further includes a language translator **900**.

The language translator **900** may be a single device that translates text data in one language to text data in multiple other selected languages or may be separate devices that translate text data between two selected languages (for example, English and Japanese).

In this embodiment, the speech recognition device **203** serves to convert the speech input signals into text data of the particular language of the speaker, as described above with reference to the first embodiment, and then further translates the text data into text data of the selected language. After the speech input signals are converted into recognized word text data, the recognized word text data is input to the language translator **900**. The language translator **900** performs language translation by sending the recognized word text data in the sending party's language to a sentence-level matching device **906** that determines the proper output text data sentence structure for the identified words from the word-level matching device **302**. The sentence-level matching device **906** retrieves information from the grammar database **907** and semantics database **908** relating to the rules of sentence structure and the meaning of the identified words in order to place the text data in the proper order. The text data is then output to the language translation device **901** which translates the recognized sentences into the language of the receiving party. Such language translation devices are generally known in the art and are described, for example, in U.S. Pat. No. 5,497,319 to Chong et al., U.S. Pat. No. 5,640,575 to Maruyama et al. and U.S. Pat. No. 5,477,450 to Takeda et al., the disclosures of which are hereby incorporated by reference.

In a fourth embodiment, the speech recognition device **203** shown in FIG. **3** converts the speech input signals into textual representations of the speech signals. In this embodiment, the textual representation is not the actual words that are part of the speech input signals but is rather a textual code representing the phonemes detected in the speech input signals.

In this embodiment, the phonemes detected from the feature vectors are identified using the word model device **303**. The word model device **303** includes a listing of phonemes and corresponding textual code representations of these phonemes. These textual code representations are then output to the outgoing text buffer **207**.

The textual code representations are then transmitted to the receiving subscriber terminal **110** in the manner set forth above. The receiving subscriber terminal **110** would further include in the text to speech conversion device **206** a phonemes database (not shown) for converting the textual code representations into the spoken phonemes. Thus, rather than having to determine the actual words in the speech input, only the phonemes are detected and sent as text data.

As shown in FIG. **1**, the method of this invention is preferably implemented on a programmed general purpose computer. However, the speech to text and text to speech conversions and transmission can also be implemented on a special purpose computer, a programmed microprocessor or microcontroller and peripheral integrated circuit elements, an ASIC or other integrated circuit, a hardware electronic or logic circuit such as a discrete element circuit, a programmable logic device such as a PLD, PLA, FPGA or PAL, or the like. In general, any device on which a finite state machine capable of implementing the flowcharts shown in FIGS. **5–8** can be used to implement the speech to text and text to speech conversions and transmission of this invention.

While this invention has been described with specific embodiments thereof, it is evident that many alternatives, modifications, and variations will be apparent to those skilled in the art. Accordingly, the preferred embodiments of the invention as set forth herein are intended to be illustrative, not limiting. Various changes may be made without departing from the spirit and scope of the invention.

What is claimed is:

1. A method of providing communication between a first party using a first communication device and a second party using a second communication device over a network, comprising:

generating a first textual representation by the first party using the first communication device;

sending the first textual representation over the network to the second communication device;

selecting a first speech pattern by the second party for converting the first textual representation into first speech output signals; and

converting the first textual representation into first speech output signals using the first selected speech pattern.

2. The method of claim **1**, wherein the first selected speech pattern is selected from a plurality of speech patterns of different types of speakers stored in the second communication device.

3. The method of claim **1**, wherein generating the first textual representation includes:

receiving, in the first communication device, first speech input from the first party; and

converting the first speech input into the first textual representation.

4. The method of claim **3**, further comprising:

receiving, by the first party, a second textual representation sent from the second party over the network; and

converting the second textual representation into second speech output signals.

5. The method of claim **4**, wherein communication between the first party and the second party is substantially continuous realtime communication.

6. The method of claim **4**, further comprising storing at least one of the first textual representation and the second textual representation in a storage device for subsequent retrieval.

7. The method of claim **4**, wherein the step of converting the second textual representation includes determining a second speech pattern to use for outputting the second speech output signals.

8. The method of claim **3**, wherein the step of converting the first speech input signals includes determining phonemes included in the first speech input signals and converting the phonemes into at least one of textual words and textual codes.

9. The method of claim **3**, wherein the step of converting the first speech input signals includes determining textual words included in the first speech input signals and arranging the textual words into textual sentences using grammar and semantic information for a first language.

10. The method of claim **9**, wherein the step of converting the first speech input signals further includes translating the textual words of the first language into textual words of a second language.

11. The method of claim **9**, wherein communication between the first party and the second part is substantially continuous realtime communication.

12. The method of claim **1**, wherein selecting a speech pattern further comprises, prior to the first speech pattern being selected by the second party, determining if another speech pattern has been designated by the first party, and if the first party has not designated another speech pattern, then selecting the first speech pattern selected by the second party.

13. A communication apparatus that provides communication between a first party and a second party over a network, comprising:

an incoming text buffer that receives a first textual representation from the first party over the network;

a text-to-speech conversion device that converts the first textual representation into first speech output signals, wherein the text-to-speech conversion device uses a first speech pattern for outputting the first speech output signals, and wherein the first speech pattern is selected by the second party; and

a speech output device that outputs the first speech output signals.

14. The communication apparatus of claim **13**, wherein the first speech pattern is selected from a plurality of speech patterns stored in the communication apparatus.

15. The communication apparatus of claim **13**, further comprising:

a speech input device that receives speech input from the second party and converts the speech input into speech input signals;

a speech-to-text device that converts the speech input signals into a second textual representation of the speech input signals; and

a communications interface that sends the second textual representation over the network to the first party.

16. The communication apparatus of claim **15**, further comprising a storage device for storing at least one of the first textual representation and the second textual representation for subsequent retrieval.

17. The communication apparatus of claim **15**, wherein the communication apparatus provides substantially continuous realtime communication between the first party and the second party.

18. The communication apparatus of claim **15**, wherein the speech-to-text device determines phonemes included in the speech input signals and converts the phonemes into at least one of textual words and textual codes.

19. The communication apparatus of claim **15**, wherein the speech-to-text device determines the textual words included in the speech input signals and arranges the textual words into textual sentences using grammar and semantic information for a first language.

20. The communication apparatus of claim **19**, wherein the speech-to-text device translates the textual words of the first language into textual words of a second language.

* * * * *