JS 44 (Rev. 07/16)

CIVIL COVER SHEET

The JS 44 civil cover sheet and the information contained herein neither replace nor supplement the filing and service of pleadings or other papers as required by law, except as provided by local rules of court. This form, approved by the Judicial Conference of the United States in September 1974, is required for the use of the Clerk of Court for the purpose of initiating the civil docket sheet. *(SEE INSTRUCTIONS ON NEXT PAGE OF THIS FORM.)*

I. (a) PLAINTIFFS

Promptu Systems Corporation

(b) County of Residence of First Listed Plaintiff San Mateo County, CA
(EXCEPT IN U.S. PLAINTIFF CASES)

(c) Attorneys *(Firm Name, Address, and Telephone Number)*
See attachment.

DEFENDANTS

Comcast Corporation and
Comcast Cable Communications, LLC

County of Residence of First Listed Defendant Philadelphia County, PA
(IN U.S. PLAINTIFF CASES ONLY)

NOTE: IN LAND CONDEMNATION CASES, USE THE LOCATION OF THE TRACT OF LAND INVOLVED.

Attorneys *(If Known)*

II. BASIS OF JURISDICTION *(Place an "X" in One Box Only)*

- ☐ 1 U.S. Government Plaintiff
- ☒ 3 Federal Question *(U.S. Government Not a Party)*
- ☐ 2 U.S. Government Defendant
- ☐ 4 Diversity *(Indicate Citizenship of Parties in Item III)*

III. CITIZENSHIP OF PRINCIPAL PARTIES *(Place an "X" in One Box for Plaintiff and One Box for Defendant)*
(For Diversity Cases Only)

	PTF	DEF		PTF	DEF
Citizen of This State	☐ 1	☐ 1	Incorporated *or* Principal Place of Business In This State	☐ 4	☐ 4
Citizen of Another State	☐ 2	☐ 2	Incorporated *and* Principal Place of Business In Another State	☐ 5	☐ 5
Citizen or Subject of a Foreign Country	☐ 3	☐ 3	Foreign Nation	☐ 6	☐ 6

IV. NATURE OF SUIT *(Place an "X" in One Box Only)*

CONTRACT
- ☐ 110 Insurance
- ☐ 120 Marine
- ☐ 130 Miller Act
- ☐ 140 Negotiable Instrument
- ☐ 150 Recovery of Overpayment & Enforcement of Judgment
- ☐ 151 Medicare Act
- ☐ 152 Recovery of Defaulted Student Loans (Excludes Veterans)
- ☐ 153 Recovery of Overpayment of Veteran's Benefits
- ☐ 160 Stockholders' Suits
- ☐ 190 Other Contract
- ☐ 195 Contract Product Liability
- ☐ 196 Franchise

REAL PROPERTY
- ☐ 210 Land Condemnation
- ☐ 220 Foreclosure
- ☐ 230 Rent Lease & Ejectment
- ☐ 240 Torts to Land
- ☐ 245 Tort Product Liability
- ☐ 290 All Other Real Property

TORTS

PERSONAL INJURY
- ☐ 310 Airplane
- ☐ 315 Airplane Product Liability
- ☐ 320 Assault, Libel & Slander
- ☐ 330 Federal Employers' Liability
- ☐ 340 Marine
- ☐ 345 Marine Product Liability
- ☐ 350 Motor Vehicle
- ☐ 355 Motor Vehicle Product Liability
- ☐ 360 Other Personal Injury
- ☐ 362 Personal Injury - Medical Malpractice

PERSONAL INJURY
- ☐ 365 Personal Injury - Product Liability
- ☐ 367 Health Care/ Pharmaceutical Personal Injury Product Liability
- ☐ 368 Asbestos Personal Injury Product Liability

PERSONAL PROPERTY
- ☐ 370 Other Fraud
- ☐ 371 Truth in Lending
- ☐ 380 Other Personal Property Damage
- ☐ 385 Property Damage Product Liability

CIVIL RIGHTS
- ☐ 440 Other Civil Rights
- ☐ 441 Voting
- ☐ 442 Employment
- ☐ 443 Housing/ Accommodations
- ☐ 445 Amer. w/Disabilities - Employment
- ☐ 446 Amer. w/Disabilities - Other
- ☐ 448 Education

PRISONER PETITIONS

Habeas Corpus:
- ☐ 463 Alien Detainee
- ☐ 510 Motions to Vacate Sentence
- ☐ 530 General
- ☐ 535 Death Penalty

Other:
- ☐ 540 Mandamus & Other
- ☐ 550 Civil Rights
- ☐ 555 Prison Condition
- ☐ 560 Civil Detainee - Conditions of Confinement

FORFEITURE/PENALTY
- ☐ 625 Drug Related Seizure of Property 21 USC 881
- ☐ 690 Other

LABOR
- ☐ 710 Fair Labor Standards Act
- ☐ 720 Labor/Management Relations
- ☐ 740 Railway Labor Act
- ☐ 751 Family and Medical Leave Act
- ☐ 790 Other Labor Litigation
- ☐ 791 Employee Retirement Income Security Act

IMMIGRATION
- ☐ 462 Naturalization Application
- ☐ 465 Other Immigration Actions

BANKRUPTCY
- ☐ 422 Appeal 28 USC 158
- ☐ 423 Withdrawal 28 USC 157

PROPERTY RIGHTS
- ☐ 820 Copyrights
- ☒ 830 Patent
- ☐ 840 Trademark

SOCIAL SECURITY
- ☐ 861 HIA (1395ff)
- ☐ 862 Black Lung (923)
- ☐ 863 DIWC/DIWW (405(g))
- ☐ 864 SSID Title XVI
- ☐ 865 RSI (405(g))

FEDERAL TAX SUITS
- ☐ 870 Taxes (U.S. Plaintiff or Defendant)
- ☐ 871 IRS—Third Party 26 USC 7609

OTHER STATUTES
- ☐ 375 False Claims Act
- ☐ 376 Qui Tam (31 USC 3729(a))
- ☐ 400 State Reapportionment
- ☐ 410 Antitrust
- ☐ 430 Banks and Banking
- ☐ 450 Commerce
- ☐ 460 Deportation
- ☐ 470 Racketeer Influenced and Corrupt Organizations
- ☐ 480 Consumer Credit
- ☐ 490 Cable/Sat TV
- ☐ 850 Securities/Commodities/ Exchange
- ☐ 890 Other Statutory Actions
- ☐ 891 Agricultural Acts
- ☐ 893 Environmental Matters
- ☐ 895 Freedom of Information Act
- ☐ 896 Arbitration
- ☐ 899 Administrative Procedure Act/Review or Appeal of Agency Decision
- ☐ 950 Constitutionality of State Statutes

V. ORIGIN *(Place an "X" in One Box Only)*

☒ 1 Original Proceeding ☐ 2 Removed from State Court ☐ 3 Remanded from Appellate Court ☐ 4 Reinstated or Reopened ☐ 5 Transferred from Another District *(specify)* ☐ 6 Multidistrict Litigation - Transfer ☐ 8 Multidistrict Litigation - Direct File

VI. CAUSE OF ACTION

Cite the U.S. Civil Statute under which you are filing *(Do not cite jurisdictional statutes unless diversity)*:
35 U.S.C. 1 et seq.

Brief description of cause:
Patent infringement and unfair uncompetition

VII. REQUESTED IN COMPLAINT:

☐ CHECK IF THIS IS A CLASS ACTION UNDER RULE 23, F.R.Cv.P.

DEMAND $

CHECK YES only if demanded in complaint:
JURY DEMAND: ☒ Yes ☐ No

VIII. RELATED CASE(S) IF ANY

(See instructions): JUDGE _____ DOCKET NUMBER _____

DATE 12/19/2016

SIGNATURE OF ATTORNEY OF RECORD

FOR OFFICE USE ONLY

RECEIPT # _____ AMOUNT _____ APPLYING IFP _____ JUDGE _____ MAG. JUDGE _____

Attorneys for Plaintiff

White and Williams LLP
David H. Marion, Esq.
Frank A. Bruno, Esq.
1650 Market Street
One Liberty Place, Ste. 1800
Philadelphia, PA 19103-7395
215.864.6870 and 215.864.6225

Carr / Ferrell, LLP
Scott R. Mosko, Esq.
Robert J. Yorio, Esq.
120 Constitution Drive
Menlo Park, California 94025
650.812.3400

IN THE UNITED STATES DISTRICT COURT
FOR THE EASTERN DISTRICT OF PENNSYLVANIA

CASE MANAGEMENT TRACK DESIGNATION FORM

Promptu Systems Corporation	:	CIVIL ACTION
v.	:	
Comcast Corporation and Comcast Cable Communications, LLC	:	NO.

In accordance with the Civil Justice Expense and Delay Reduction Plan of this court, counsel for plaintiff shall complete a Case Management Track Designation Form in all civil cases at the time of filing the complaint and serve a copy on all defendants. (See § 1:03 of the plan set forth on the reverse side of this form.) In the event that a defendant does not agree with the plaintiff regarding said designation, that defendant shall, with its first appearance, submit to the clerk of court and serve on the plaintiff and all other parties, a Case Management Track Designation Form specifying the track to which that defendant believes the case should be assigned.

SELECT ONE OF THE FOLLOWING CASE MANAGEMENT TRACKS:

(a) Habeas Corpus – Cases brought under 28 U.S.C. § 2241 through § 2255. ( )

(b) Social Security – Cases requesting review of a decision of the Secretary of Health and Human Services denying plaintiff Social Security Benefits. ( )

(c) Arbitration – Cases required to be designated for arbitration under Local Civil Rule 53.2. ( )

(d) Asbestos – Cases involving claims for personal injury or property damage from exposure to asbestos. ( )

(e) Special Management – Cases that do not fall into tracks (a) through (d) that are commonly referred to as complex and that need special or intense management by the court. (See reverse side of this form for a detailed explanation of special management cases.) (☑)

(f) Standard Management – Cases that do not fall into any one of the other tracks. ( )

12/19/2016	David H. Marion	Promptu Systems Corporation
Date	**Attorney-at-law**	**Attorney for**
215-864-6870	215-789-6675	mariond@whiteandwilliams.com
Telephone	**FAX Number**	**E-Mail Address**

(Civ. 660) 10/02

UNITED STATES DISTRICT COURT

FOR THE EASTERN DISTRICT OF PENNSYLVANIA — DESIGNATION FORM to be used by counsel to indicate the category of the case for the purpose of assignment to appropriate calendar.

Address of Plaintiff: Promptu Systems Corporation, 333 Ravenswood Avenue, Building 201, Menlo Park, CA

Address of Defendant: Comcast Corp./Comcast Cable Communications, LLC, One Comcast Center, 1701 JFK Blvd., Philadelphia PA

Place of Accident, Incident or Transaction: Nationwide

(*Use Reverse Side For Additional Space*)

Does this civil action involve a nongovernmental corporate party with any parent corporation and any publicly held corporation owning 10% or more of its stock?

(Attach two copies of the Disclosure Statement Form in accordance with Fed.R.Civ.P. 7.1(a)) Yes☐ No☑

Does this case involve multidistrict litigation possibilities? Yes☐ No☑

RELATED CASE, IF ANY:

Case Number: _____ Judge _____ Date Terminated: _____

Civil cases are deemed related when yes is answered to any of the following questions:

1. Is this case related to property included in an earlier numbered suit pending or within one year previously terminated action in this court? Yes☐ No☑
2. Does this case involve the same issue of fact or grow out of the same transaction as a prior suit pending or within one year previously terminated action in this court? Yes☐ No☑
3. Does this case involve the validity or infringement of a patent already in suit or any earlier numbered case pending or within one year previously terminated action in this court? Yes☐ No☑
4. Is this case a second or successive habeas corpus, social security appeal, or pro se civil rights case filed by the same individual? Yes☐ No☑

CIVIL: (Place ✔ in ONE CATEGORY ONLY)

A. *Federal Question Cases*:

1. ☐ Indemnity Contract, Marine Contract, and All Other Contracts
2. ☐ FELA
3. ☐ Jones Act-Personal Injury
4. ☐ Antitrust
5. ☑ Patent
6. ☐ Labor-Management Relations
7. ☐ Civil Rights
8. ☐ Habeas Corpus
9. ☐ Securities Act(s) Cases
10. ☐ Social Security Review Cases
11. ☐ All other Federal Question Cases

(Please specify) _____

B. *Diversity Jurisdiction Cases*:

1. ☐ Insurance Contract and Other Contracts
2. ☐ Airplane Personal Injury
3. ☐ Assault, Defamation
4. ☐ Marine Personal Injury
5. ☐ Motor Vehicle Personal Injury
6. ☐ Other Personal Injury (Please specify)
7. ☐ Products Liability
8. ☐ Products Liability — Asbestos
9. ☐ All other Diversity Cases

(Please specify) _____

ARBITRATION CERTIFICATION

(*Check Appropriate Category*)

I, David H. Marion, counsel of record do hereby certify:

☑ Pursuant to Local Civil Rule 53.2, Section 3(c)(2), that to the best of my knowledge and belief, the damages recoverable in this civil action case exceed the sum of $150,000.00 exclusive of interest and costs;

☑ Relief other than monetary damages is sought.

DATE: 12/19/2016 _____ Attorney-at-Law 3590 Attorney I.D.#

NOTE: A trial de novo will be a trial by jury only if there has been compliance with F.R.C.P. 38.

I certify that, to my knowledge, the within case is not related to any case now pending or within one year previously terminated action in this court except as noted above.

DATE: 12/19/2016 _____ Attorney-at-Law 3590 Attorney I.D.#

CIV. 609 (5/2012)

UNITED STATES DISTRICT COURT

FOR THE EASTERN DISTRICT OF PENNSYLVANIA — DESIGNATION FORM to be used by counsel to indicate the category of the case for the purpose of assignment to appropriate calendar.

Address of Plaintiff: Promptu Systems Corporation, 333 Ravenswood Avenue, Building 201, Menlo Park, CA

Address of Defendant: Comcast Corp./Comcast Cable Communications, LLC, One Comcast Center, 1701 JFK Blvd., Philadelphia PA

Place of Accident, Incident or Transaction: Nationwide

(Use Reverse Side For Additional Space)

Does this civil action involve a nongovernmental corporate party with any parent corporation and any publicly held corporation owning 10% or more of its stock?

(Attach two copies of the Disclosure Statement Form in accordance with Fed.R.Civ.P. 7.1(a)) Yes☐ No☑

Does this case involve multidistrict litigation possibilities? Yes☐ No☑

RELATED CASE, IF ANY:

Case Number: _____ Judge _____ Date Terminated: _____

Civil cases are deemed related when yes is answered to any of the following questions:

1. Is this case related to property included in an earlier numbered suit pending or within one year previously terminated action in this court? Yes☐ No☑
2. Does this case involve the same issue of fact or grow out of the same transaction as a prior suit pending or within one year previously terminated action in this court? Yes☐ No☑
3. Does this case involve the validity or infringement of a patent already in suit or any earlier numbered case pending or within one year previously terminated action in this court? Yes☐ No☑
4. Is this case a second or successive habeas corpus, social security appeal, or pro se civil rights case filed by the same individual? Yes☐ No☑

CIVIL: (Place ✔ in ONE CATEGORY ONLY)

A. *Federal Question Cases*:

1. ☐ Indemnity Contract, Marine Contract, and All Other Contracts
2. ☐ FELA
3. ☐ Jones Act-Personal Injury
4. ☐ Antitrust
5. ☑ Patent
6. ☐ Labor-Management Relations
7. ☐ Civil Rights
8. ☐ Habeas Corpus
9. ☐ Securities Act(s) Cases
10. ☐ Social Security Review Cases
11. ☐ All other Federal Question Cases
    (Please specify) _____

B. *Diversity Jurisdiction Cases*:

1. ☐ Insurance Contract and Other Contracts
2. ☐ Airplane Personal Injury
3. ☐ Assault, Defamation
4. ☐ Marine Personal Injury
5. ☐ Motor Vehicle Personal Injury
6. ☐ Other Personal Injury (Please specify)
7. ☐ Products Liability
8. ☐ Products Liability — Asbestos
9. ☐ All other Diversity Cases
   (Please specify) _____

ARBITRATION CERTIFICATION

(Check Appropriate Category)

I, David H. Marion, counsel of record do hereby certify:

☑ Pursuant to Local Civil Rule 53.2, Section 3(c)(2), that to the best of my knowledge and belief, the damages recoverable in this civil action case exceed the sum of $150,000.00 exclusive of interest and costs;

☑ Relief other than monetary damages is sought.

DATE: 12/19/2016 _____ Attorney-at-Law 3590 Attorney I.D.#

NOTE: A trial de novo will be a trial by jury only if there has been compliance with F.R.C.P. 38.

I certify that, to my knowledge, the within case is not related to any case now pending or within one year previously terminated action in this court except as noted above.

DATE: 12/19/2016 _____ Attorney-at-Law 3590 Attorney I.D.#

CIV. 609 (5/2012)

IN THE UNITED STATES DISTRICT COURT
FOR THE EASTERN DISTRICT OF PENNSYLVANIA

PROMPTU SYSTEMS CORPORATION Plaintiff, v. COMCAST CORPORATION and COMCAST CABLE COMMUNICATIONS, LLC Defendants.	CIVIL ACTION No. _____ **RULE 7.1 DISCLOSURE STATEMENT**

The nongovernmental corporate party, Promptu Systems Corporation in the above-listed civil action does not have any parent corporation or publicly held corporation that owns 10% or more of its stock.

WHITE AND WILLIAMS LLP

BY: _____
David H. Marion

WHITE AND WILLIAMS LLP
David H. Marion (I.D. No. 3590)
1650 Market Street
One Liberty Place, Suite 1800
Philadelphia, PA 19103-7395
Tel. (215) 864-7000

CARR AND FERRELL LLP
Scott R. Mosko *(pro hac vice* pending)
Robert J. Yorio *(pro hac vice* pending)
120 Constitution Drive
Menlo Park, CA 94025
Tel. (650) 812-3400

Dated: December 19, 2016

IN THE UNITED STATES DISTRICT COURT
FOR THE EASTERN DISTRICT OF PENNSYLVANIA

PROMPTU SYSTEMS CORPORATION Plaintiff, v. COMCAST CORPORATION and COMCAST CABLE COMMUNICATIONS, LLC Defendants.	CIVIL ACTION No. _____ **JURY TRIAL DEMANDED**

CIVIL ACTION COMPLAINT WITH JURY TRIAL DEMAND

SUMMARY OF ACTION

1. Plaintiff Promptu Systems Corporation ("Promptu") brings this action against Comcast Corporation and Comcast Cable Communications, LLC (separately referred to herein as "Comcast Cable" and together with Comcast Corporation referred to herein as "Comcast") as follows:

a. First, for damages and injunctive relief by reason of Defendants' infringement of three United States patents assigned to Promptu, specifically U.S. Patent Nos. 7,260,538 (the "'538 Patent"); 7,047,196 (the "'196 Patent"); and RE44,326 (the "'326 Patent") (collectively referred to as the "Patents-in-Suit"); and

b. Second, for compensatory and punitive damages for claims of common law unfair competition, breaches of implied-in-law and implied-in-fact contract and unjust enrichment arising from Defendants' entering into a confidential relationship with Promptu designed to elicit its know-how, and then wrongfully using this information to deploy a voice

recognition system in Comcast's XFINITY® and X1 Entertainment Operating Systems® without compensating Promptu.

THE PARTIES

2. Promptu is a Delaware corporation with a principal place of business at 333 Ravenswood Avenue, Building 201, Menlo Park, California.

3. On information and belief, Comcast Corporation is a Pennsylvania corporation with a principal place of business at One Comcast Center, 1701 John F. Kennedy Blvd., Philadelphia, Pennsylvania 19103.

4. On information and belief, Comcast Cable Communications, LLC is a Delaware limited liability company with a principal place of business at One Comcast Center, 1701 John F. Kennedy Blvd., Philadelphia, Pennsylvania 19103. Comcast Cable Communications, LLC is engaged in developing, managing and operating broadband cable networks and is indirectly wholly owned by Comcast Corporation.

JURISDICTION AND VENUE

5. This action includes claims for patent infringement concerning Comcast's manufacture, use, sale, offer for sale and/or importation in the United States of devices, including remotes, set-top boxes, receivers and others that receive and enable processing of voice commands that in turn generate content on a Comcast cable subscriber's television or other content-receiving device. This action also includes claims for patent infringement concerning Comcast's products, methods and systems as they relate to voice recognition technology implemented in services and content offered by Comcast in its XFINITY® and its X1 Entertainment Operating Systems®. As such, this action arises under the patent laws of the United States, 35 U.S.C. §§ 1 et seq. Accordingly, this Court has subject matter jurisdiction

pursuant to 28 U.S.C. §§ 1331 and 1338(a). For the non-patent claims asserted herein, this Court has jurisdiction under 28 U.S.C. § 1338(b) as a result of their considerable overlap with the federal claims.

6. Venue is proper in this district under 28 U.S.C. §§ 1391 and 1400(b).

7. This Court has general and/or specific personal jurisdiction over Comcast Corporation, and venue is proper because, on information and belief, Comcast Corporation's principal place of business is located in this district. In addition, Comcast Corporation, directly and/or in combination with its subsidiaries and/or through its agents or others related to it, does continuous and systematic business in this district including by providing infringing products and services to residents of this district, by providing infringing products and services that it knew would be used within this district, and/or by participating in the solicitation of business from residents of this district. In addition, on information and belief, Comcast Corporation, directly or through its subsidiaries, places infringing products within the stream of commerce, which are directed at this district, with the knowledge and/or understanding that such products will be sold, leased, or otherwise provided to customers within this district. In addition, on information and belief, Comcast Corporation, directly or through its subsidiaries, employs individuals within this district, including employees who provide infringing products and services to customers here, and maintains offices and facilities here. Comcast Corporation, directly or through its subsidiaries or others related to it, operates highly commercial websites through which regular sales and/or leases of products and/or sales of services are made to customers in this district, including products and services that, on information and belief, infringe the patents asserted herein and/or is liable for its wrongful activities as herein alleged.

8. This Court has general and/or specific personal jurisdiction over Comcast Cable, and venue is proper because, on information and belief, Comcast Cable's principal place of business is located in this district. In addition, Comcast Cable, directly and/or in combination with its parent, subsidiaries and/or through its agents or others related to it, does continuous and systematic business in this district including by providing infringing products and services to residents of this district, by providing infringing products and services that it knew would be used within this district, and/or by participating in the solicitation of business from residents of this district. In addition, on information and belief, Comcast Cable, directly or through its subsidiaries or others related to it, places infringing products within the stream of commerce, which are directed at this district, with the knowledge and/or understanding that such products will be sold, leased, or otherwise provided to customers within this district. In addition, on information and belief, Comcast Cable, directly or through its subsidiaries or others related to it, employs individuals within this district, including employees who provide infringing products and services to customers here, and maintains offices and facilities here. Comcast Cable, directly or through its subsidiaries or others related to it, operates highly commercial websites through which regular sales and/or leases of products and/or sales of services are made to customers in this district, including products and services that, on information and belief, infringe the patents asserted herein and/or is liable for its wrongful activities as herein alleged.

FACTS

9. In or about 2000, Promptu's predecessor, AgileTV Corporation ("AgileTV") was created and soon became an innovator in voice recognition and natural language understanding technology. Shortly after its inception, AgileTV further developed and applied this technology to the television and cable network industries.

10. In or about August, 2006, AgileTV changed its name to Promptu Systems Corporation. Throughout this Complaint, references to "Promptu" include AgileTV.

11. The United States Patent and Trademark Office has granted Promptu's engineers and inventors numerous patents, many of which reflect inventions applicable to voice recognition in the television and cable industries. Promptu is the assignee of multiple patents concerning voice recognition and its application to cable television, including the '538 Patent, entitled "Method and Apparatus for Voice Control of a Television Control Device," and including the '196 and '326 Patents, both entitled "System and Method of Voice Recognition Near a Wireline Node of a Network Supporting Cable Television and/or Video Delivery."

12. Comcast has published that, through its operations of Comcast Cable, it is one of the nation's largest providers of video, high-speed Internet, and voice services to residential customers. Comcast either by itself or through its wholly-owned subsidiaries or others, offers a voice recognition feature through its television cable service by and through its brand, XFINITY® and its X1 Entertainment Operating Systems®.

13. A customer who subscribes with Comcast and/or its subsidiaries or others to receive cable television network access may now receive an XFINITY®-branded system capable of receiving voice commands, and other devices enabling access to Comcast's cable system. By and through these devices, and the signals generated from and to these devices, Comcast receives and processes voice commands and provides content consistent with the spoken voice to the customer's television or other device that displays content.

14. Beginning in or about 2001, Comcast expressed an interest in Promptu's capability of implementing a voice recognition feature into Comcast's television cable network

system. Prior to this time, Comcast had no ability to adopt and implement voice recognition technology into its cable network without Promptu's expertise.

15. Consistent with and further to this expressed interest, Comcast entered into a confidential relationship with Promptu. During the course of this relationship, Promptu agreed to disclose its expertise and know-how of implementing voice recognition technology into Comcast's cable network system. Comcast encouraged these disclosures with assurances that Comcast would adopt this technology throughout its cable system and appropriately compensate Promptu.

16. The relationship of trust and confidence created by Comcast and the course of dealings between the parties lulled Promptu into the reasonable belief that it could confide in and trust Comcast with the detailed disclosures Promptu provided. Promptu thereafter disclosed its technical know-how to Comcast and enabled voice recognition on portions of Comcast's cable television network. Promptu reasonably expected that if Comcast fully deployed voice recognition into its cable network, Comcast would compensate Promptu for sharing its technology and know-how with Comcast. Comcast has now fully deployed voice recognition into its cable network, yet, Comcast has failed to compensate Promptu for disclosing its know-how, which enabled this deployment.

17. Shortly after this above-described confidential relationship was created, to elicit full disclosure of Promptu's technical knowledge and capabilities, Comcast conducted many in-person meetings with Promptu both at Comcast Headquarters in Philadelphia, Pennsylvania, and at Promptu's office in California. Several of Comcast's senior technical officials attended these meetings, including Comcast's Chief Technical Officer, Brad Dusto; Vice President of the

Digital TV Division, Mark Hess; Senior Vice President of Strategic Planning, Mark Coblitz; and Chief Financial Officer, Joe Donnolly.

18. During these initial meetings, Comcast sought and received details, comprehensive presentations and demonstrations of Promptu's voice recognition technology applied to television and cable networks.

19. At all relevant times, Comcast expressed great interest in implementing Promptu's technology into its network.

20. In addition to encouraging Promptu's full disclosure of its technical know-how, Comcast also elicited confidential business information from Promptu, including details of Promptu's financial position, projected costs to implement voice recognition technology into the Comcast network, and profitability projections.

21. Additional details elicited by Comcast and disclosed by Promptu included Promptu's intellectual property position and strategies related to its intellectual property holdings. Consistent with Comcast's request, Promptu disclosed each of its granted patents, each of its patent applications, the status of said patent applications before the United States Patent and Trademark Office and future intellectual property strategies. These disclosures, requested by Comcast and provided by Promptu included the applications that matured into the Patents-in-Suit.

22. Comcast lead Promptu to believe that it intended to move quickly to implement Promptu's technology into the Comcast network. Comcast requested and Promptu provided various economic models, cash flow analyses and revenue projections regarding the implementation of this technology into the Comcast cable network.

23. Promptu also provided to Comcast detailed writings describing how voice recognition can be implemented into the Comcast network.

24. Comcast praised Promptu upon receipt of the above-referenced presentations and information. Consistent with Comcast's goal of fully eliciting technical disclosures from Promptu, Comcast invited Promptu to make multiple presentations of its technology to senior management, including those to Comcast's President, Steve Burke and its founder Ralph Roberts. Mr. Burke encouraged Promptu to continue its work with Comcast by telling Promptu, among other things, that Comcast was in need of this technology as it would ease the problems the Comcast subscribers were having and increase the value of its Video-On-Demand products.

25. Comcast also told Promptu that implementation of Promptu's technology would further Comcast's efforts to reduce what the industry refers to as "churn" and would increase the popularity of cable television.

26. On or about May 6, 2003, and by addendum on February 9, 2004, Comcast's Senior Vice President of Strategic Planning, Mark Coblitz, signed a Memorandum of Understanding ("MOU") with Promptu. Comcast had represented that it would also invest $2 million of capital into Promptu that would allow Promptu to continue its efforts of implementing its technology into the Comcast cable network, and those earlier representations were confirmed in the MOU. This MOU stated in part, "Comcast believes that [Promptu's] speech recognition technology enabling the control and navigation of cable services on the television via voice commands has the potential to be an attractive service that could have a significant benefit to cable subscribers." Sam Schwartz, the Managing Director of Comcast Interactive Capital, the entity responsible for effecting this stated investment, suggested to Promptu that Comcast would move quickly to finalize this investment.

27. This above-referenced MOU and additional representations by Comcast reflected Comcast's intent to implement a phased deployment of Promptu's technology in the Comcast cable network. Mr. Coblitz represented to Promptu that the initial deployment would involve approximately 50,000 subscribers.

28. At Comcast's request, Promptu installed voice recognition into various locations throughout its network, including various "Labs" within Comcast that would allow its technicians full access to Promptu's technology and know-how, and the "Comcast Living Room," that would provide Comcast Executives access to, and ability to demonstrate Promptu's voice recognition technology on the Comcast cable network.

29. In addition, Comcast requested and Promptu installed its voice recognition technology into its President, Brian Robert's office so he could use and demonstrate to others this voice recognition function on the Comcast network.

30. As Promptu worked toward a viable cost effective system that could be fully deployed into the Comcast network, Promptu shared cost projections with Comcast. Comcast repeatedly assured Promptu that its proposals and expected costs were acceptable, which assurances further encouraged Promptu to reasonably believe it would be compensated after voice recognition was fully deployed into Comcast's cable network.

31. Shortly after voice recognition technology was deployed into the Comcast "Labs" and "Living Room," and consistent with the above-referenced MOU, Comcast committed to invest $2 million into Promptu. This investment further validated Promptu's expectation that it would be compensated for its disclosures and implementations once voice recognition was fully deployed in the Comcast cable network. Upon receiving this confirmation of investment, Comcast's Executive Vice President of Business, Stephan Silva told Promptu that Comcast

intended to ubiquitously deploy the voice recognition technology to its 21,000,000 subscribers so that Comcast would be the first large television content provider offering this feature. Comcast represented to Promptu that it intended to make voice recognition part of its core technology offering.

32. In or about 2004, Comcast began negotiating what would become various written agreements with Promptu which were executed by Comcast on May 11, 2005. One of these agreements, entitled "Marketing Trial Agreement for Voice Activated Television Control Service," sought Promptu's assistance in " . . . conduct[ing] a field trial of limited scope and duration in which [Comcast] . . . [would] evaluate[] [Promptu's] voice-activated remote control and programming interface products and services . . . and measure[] consumer demand for voice-activated products and services in general…"

33. Another written agreement Comcast entered into with Promptu was entitled, "License and Development Agreement." This Agreement included: (a) the details of Comcast's investment into Promptu; (b) an acquisition by Comcast of an option to license Promptu's earlier-disclosed intellectual property; (c) a process whereby Promptu would fully implement its voice-recognition technology capabilities into the Comcast cable network; and (d) a payment methodology that would apply if the contingencies set forth in the Agreement occurred.

34. Pursuant to this License and Development Agreement, Comcast proposed that Promptu create a plan that would deploy voice recognition in Comcast's cable network in phases. Consistent with this proposal, Promptu prepared a series of documents referred to as "Statements of Work" detailing a phased deployment. In or about October 2005, Comcast's Executive Vice President of Business Development, Stephan Silva told Promptu that the then-

proposed Statement of Work prepared by Promptu reflected the detail Comcast needed to fully implement voice recognition into its network.

35. Promptu was induced to believe that it would be appropriately compensated after voice recognition technology was fully deployed into the Comcast network. This belief was encouraged and justified, inter alia, by (a) Comcast's intense interest in Promptu's capabilities of implementing voice recognition into the Comcast network; (b) the Comcast executive's continuing interaction with, and encouragement of Promptu; (c) the multiple meetings between Promptu and Comcast; (d) the substantial time and effort Comcast required Promptu to expend in the course of the confidential relationship; (e) the written agreements Comcast entered into with Promptu; and (f) Comcast's favorable response to the Statement of Work prepared by Promptu for the phased deployment of Promptu's System into the entire Comcast network, all described above.

36. Contrary to these above-referenced representations, acts and inducements by Comcast that lead Promptu to reasonably expect compensation for disclosing to Comcast details of voice recognition implementation over the Comcast network, in or about 2006, Comcast terminated the voice recognition implementation by Promptu. As a result, Promptu returned the money provided by Comcast, plus interest.

37. In or about September, 2015, Comcast announced that it would begin to offer a voice recognition technology system substantially similar to that which Promptu had earlier disclosed and explained to Comcast in the course of the above-detailed confidential relationship.

FIRST CLAIM FOR RELIEF
(Infringement of the '538 Patent)

38. Promptu incorporates by reference the allegations set forth in paragraphs 1 through 37 of this Complaint.

39. The '538 Patent, entitled "Method and Apparatus for Voice Control of a Television Control Device," is attached to this Complaint as Exhibit A and is hereby incorporated by this reference.

40. Comcast has infringed one or more claims of the '538 Patent by making, having made, designing, using, distributing, importing, offering for sale and/or selling XFINITY® and the X1 Entertainment Operating System®. On information and belief, the XFINITY® and the X1 Entertainment Operating System® incorporate a platform to access a library of live and on-demand video through voice commands. The remote control, set-top box and other equipment incorporated into this XFINITY® and the X1 Entertainment Operating System® and the methods employed by Comcast to effect voice recognition on its cable network allow subscribers to access different content on the X1 platform with their voice.

41. The methods and systems practiced by Comcast through its XFINITY® and the X1 Entertainment Operating System® and the operation of the apparatus, including the units referenced above, utilized by Comcast in its XFINITY® and the X1 Entertainment Operating System® meet each limitation in several claims of the '538 Patent, including, for example, Claim 1 of this patent.

42. Comcast infringes Claim 1, a method claim for at least the following reasons:

a. Comcast admits in its publications that the architecture upon which it's XFINITY® and the X1 Entertainment Operating System® cable network system is built includes a "head-end" where "content is aggregated and disseminated," and,

b. Comcast admits in its publications that its devices include a "remote" that allows the subscriber to "use simple voice commands to find something to watch easily," and

c. Comcast admits in its directions for setting up its devices in the XFINITY® and the X1 Entertainment Operating System® that the remote is paired with a "set-top box" "for control of the TV."

43. In addition to Claim 1, Comcast, through its XFINITY® and the X1 Entertainment Operating System® infringes at least the following claims of the '538 Patent: 2, 3, 4, 6, 18, 19, 20, 21, 22, 23, 24, 33, 34, 35, 40, and 41.

44. Comcast's conduct is willful and deliberate.

45. As a direct and proximate result of Comcast's acts of patent infringement, Promptu has been and continues to be injured and has sustained, and will continue to sustain substantial damages in an amount not yet determined.

SECOND CLAIM FOR RELIEF
(Infringement of the '196 Patent)

46. Promptu incorporates by reference the allegations set forth in paragraphs 1 through 45 of this Complaint.

47. The '196 Patent, entitled "System and Method of Voice Recognition Near a Wireline Node of a Network Supporting Cable Television and/or Video Delivery" is attached to this Complaint as Exhibit B and is hereby incorporated by this reference.

48. Comcast has infringed one or more claims of the '196 Patent by making, having made, designing, using, distributing, importing, offering for sale and/or selling XFINITY® and the X1 Entertainment Operating System®. On information and belief, the XFINITY® and the X1 Entertainment Operating System® incorporate a platform to access a library of live and on-demand video through voice commands. The remote control, set-top box and other equipment incorporated into this XFINITY® and the X1 Entertainment Operating System® and the methods

employed by Comcast to effect voice recognition on its cable network allow subscribers to access different content on the X1 platform with their voice.

49. The methods and systems practiced by Comcast through its XFINITY® and the X1 Entertainment Operating System® and the operation of the apparatus utilized by Comcast in its XFINITY® and the X1 Entertainment Operating System® meet each limitation in several claims of the '196 Patent, including, for example, Claim 2 of this patent.

50. Comcast infringes Claim 2, a method claim for at least the following reasons:

a. Comcast admits that its XFINITY® and the X1 Entertainment Operating System® use "fiber optic cable to send and receive signals to/from nodes located in neighborhoods where forward (or downstream) signals are transferred from light, on the fiber optic infrastructure, to Radio Frequencies (RF) for insertion onto the coaxial cable infrastructure for ultimate transmission and distribution to residents and business served by the particular node …", and

b. Comcast admits that these nodes "receive RF signals from subscribers via the coaxial cable infrastructure which feeds these signals into the node. The node converts these signals to light for transmissions, via the fiber optic infrastructure, to Comcast's headend or hub. Some return signals include upstream data, ordering information for [video on demand] and other video services …", and

c. Comcast admits that the remote that operates the XFINITY® and the X1 Entertainment Operating System® is paired with a TV box, and once paired with a TV box, the remote will not control any other TV box.

51. In addition to Claim 2, Comcast, through its XFINITY® and the X1 Entertainment Operating System® infringes at least the following claims of the '196 Patent: 4, 15, 54, and 66.

52. Comcast's conduct is willful and deliberate.

53. As a direct and proximate result of Comcast's acts of patent infringement, Promptu has been and continues to be injured and has sustained, and will continue to sustain substantial damages in an amount not yet determined.

THIRD CLAIM FOR RELIEF
(Infringement of the '326 Patent)

54. Promptu incorporates by reference the allegations set forth in paragraphs 1 through 53 of this Complaint.

55. The '326 Patent, entitled "System and Method of Voice Recognition Near a Wireline Node of a Network Supporting Cable Television and/or Video Delivery" is attached to this Complaint as Exhibit C and is hereby incorporated by this reference.

56. Comcast has infringed one or more claims of the '326 Patent by making, having made, designing, using, distributing, importing, offering for sale and/or selling XFINITY® and the X1 Entertainment Operating System®. The XFINITY® and the X1 Entertainment Operating System® incorporate a platform to access a library of live and on-demand video through voice commands. The remote control, set-top box and other equipment incorporated into the Comcast system allow subscribers to access different content on the X1 platform with their voice.

57. The methods, systems and apparatus in the XFINITY® and the X1 Entertainment Operating System® infringe various claims of the '326 Patent, as found, for example in Claim 11 of this patent.

58. Comcast infringes Claim 11, a method claim for at least the following reasons:

a. Comcast admits that after a subscriber speaks into the remote for its XFINITY® and the X1 Entertainment Operating System®, "the voice commands are sent to Comcast and its contracted service provider for processing. Comcast and its provider use these

voice commands to provide the voice control service…", and

b. Comcast admits that "[o]nce the ASR module processes the user's request, the output (i.e. the request in text form) is sent on to the NLP [(Natural Language Processing)] module. NLP starts by parsing the text and uses a resulting parse tree to interpret the utterance, inferring the semantics from the rules used and the tags assigned in the parsing process. Once the NLP module has identified all the pieces of the request, it assembles them into a request structure that is passed to the Action Recognition (AR) module."

59. Comcast's conduct is willful and deliberate.

60. As a direct and proximate result of Comcast's acts of patent infringement, Promptu has been and continues to be injured and has sustained, and will continue to sustain substantial damages in an amount not yet determined.

FOURTH CLAIM FOR RELIEF
(Common Law Unfair Competition)

61. Promptu incorporates by reference the allegations set forth in paragraphs 1 through 60 of this Complaint.

62. Comcast entered into a confidential relationship with Promptu for the purpose of implementing Promptu's voice recognition technology into its cable network based on its above-alleged acts, its representations to Promptu, its course of dealing with Promptu, and the written agreements it signed with Promptu.

63. In the course of this confidential relationship, Comcast induced Promptu to disclose its valuable and confidential information, including the following:

a. Promptu's expertise and technical know-how regarding the implementation of its voice recognition technology into Comcast's cable network system,

b. Promptu's financial position, projected costs to implement voice recognition technology into the Comcast network, and profitability projections, and

c. Promptu's intellectual property position, including each of its granted and pending patents, and strategies related to its intellectual property holdings.

64. Based on this confidential relationship and the representations and promises by Comcast, Promptu was lulled into the belief that it could trust and confide in Comcast with the disclosures Promptu made to Comcast.

65. Based on this confidential relationship and the representations and promises by Comcast, Promptu reasonably believed that Comcast would adequately compensate Promptu upon Comcast ubiquitously adopting the disclosed details regarding voice recognition into its cable network.

66. Comcast did adopt and use, and is currently using, some or all of the disclosed information from Promptu made during the course of this above-described confidential relationship, but has not compensated Promptu.

67. Promptu has been damaged in a yet undetermined amount as a result of Comcast's acts of unfair competition as alleged herein.

FIFTH CLAIM FOR RELIEF
(Breach of Implied-in-Fact Contract)

68. Promptu incorporates by reference the allegations set forth in paragraphs 1 through 67 of this Complaint.

69. Comcast impliedly promised to compensate Promptu for disclosing, sharing and implementing voice recognition technology on its cable network as demonstrated by at least the following reasons:

a. Before Promptu shared its know-how of implementing voice recognition onto Comcast's cable network, Comcast had no ability to implement voice recognition through any other means,

b. Comcast entered into a confidential relationship with Promptu the purpose of which was to elicit Promptu's know-how and capabilities for implementing voice recognition into Comcast's cable network,

c. Comcast encouraged Promptu to disclose its know-how and capabilities of implementing voice recognition into Comcast's cable network by arranging for and attending numerous meetings between Promptu and Comcast key executives and technology experts who emphasized the importance and goals of adding voice recognition to its network,

d. Comcast entered into written agreements with Promptu expressing its goal of adding voice recognition to its cable network, including a License and Development Agreement which confirmed that if Comcast ubiquitously deployed voice recognition over its cable network, Promptu would be fully compensated,

e. Comcast officials approved the writings Promptu provided as called for in the above-described License and Development Agreement that would have obligated Comcast to compensate Promptu, but shortly thereafter ended the relationship between the parties, and

f. Without notice to Promptu, Comcast thereafter implemented voice recognition on its cable network by using part or all of the disclosed information received from Promptu.

70. Notwithstanding Comcast's termination of the relationship, the above-alleged facts and Comcast's course of conduct and dealing demonstrate Comcast impliedly represented it would adequately compensate Promptu for its efforts, expertise, and work in revealing for

Comcast's benefit the details of implementing voice recognition into Comcast's cable network if and when Comcast ubiquitously deployed this technology in its network.

71. Promptu has been damaged in a yet undetermined amount as a result of Comcast's breach of an implied-in-fact contract, as alleged herein.

SIXTH CLAIM FOR RELIEF
(Breach of Implied-in-Law Contract)

72. Promptu incorporates by reference the allegations set forth in paragraphs 1 through 71 of this Complaint.

73. Promptu has conferred valuable benefits upon Comcast by disclosing its technical know-how of implementing Promptu's voice recognition technology into Comcast's cable network at a time when Comcast had no ability to implement voice recognition through any other means.

74. Comcast adopted many of the disclosures Promptu provided to it in the launching and maintaining a voice recognition feature in its cable network, by and through its brand, XFINITY® and its X1 Entertainment Operating Systems®.

75. Comcast failed to compensate Promptu for these disclosures.

76. Promptu has been damaged in a yet undetermined amount as a result of Comcast's breach of an implied-in-law contract, as alleged herein.

SEVENTH CLAIM FOR RELIEF
(Unjust Enrichment)

77. Promptu incorporates by reference the allegations set forth in paragraphs 1 through 76 of this Complaint.

78. Promptu has conferred valuable benefits upon Comcast by disclosing its technical know-how of implementing Promptu's voice recognition technology into Comcast's cable

network at a time when Comcast had no ability to implement voice recognition through any other means.

79. Comcast adopted many of the disclosures Promptu provided to it in the launching and maintaining a voice recognition feature in its cable network.

80. Comcast has failed to compensate Promptu for these disclosures.

81. Comcast has been unjustly enriched in an undetermined amount by obtaining and using the information it obtained from Promptu and benefitting thereby.

PRAYER FOR RELIEF

WHEREFORE, Plaintiff respectfully requests the following relief:

a) An entry of judgment that each of the Patents-in-Suit is valid and enforceable;

b) An entry of judgment that Defendants have infringed one or more claims of the Patents-in-Suit;

c) An award of damages adequate to compensate Plaintiff for Defendants' infringement of the Patents-in-Suit together with prejudgment and post-judgment interest and costs, in an amount according to proof;

d) An entry of permanent injunction enjoining Defendants and their respective officers, agents, employees, and those acting in privity with them, from further infringement of the Patents-in-Suit, and further use of the confidential information and methodologies Defendants obtained from Plaintiff unless and until appropriate compensation is provided to Plaintiff;

e) An entry of judgment awarding Plaintiff all damages, including treble damages, based on any infringement found to be willful, pursuant to 35 U.S.C. § 284, together with prejudgment interest;

f) An entry of judgment of actual and punitive damages suffered by Plaintiff as a result of Defendants' unlawful conduct, in an amount to be proven at trial, as well as prejudgment interest as authorized by law;

g) An entry of judgment that this is an exceptional case and an award to Plaintiff of its costs and reasonable attorneys' fees incurred in this action as provided by 35 U.S.C. § 285;

h) An award to Plaintiff of Defendants' profits and gains resulting from Defendants' willful acts of unfair competition and/or Defendants' breach of implied-in-fact or implied-in-law contract;

i) An award to Plaintiff, in law or equity, of such other costs, expenses, and further relief as the Court may deem just and proper.

DEMAND FOR JURY TRIAL

Pursuant to Rule 38(b) of the Federal Rules of Civil Procedure, Plaintiff hereby demands trial by jury on all issues raised by the Complaint.

Dated: 12/19/16

WHITE AND WILLIAMS LLP

By _____
David H. Marion

WHITE AND WILLIAMS LLP
David H. Marion (I.D. No. 3590)
1650 Market Street
One Liberty Place, Suite 1800
Philadelphia, PA 19103-7395
Tel. (215) 864-7000

CARR AND FERRELL LLP
Scott R. Mosko (*pro hac vice* pending)
Robert J. Yorio (*pro hac vice* pending)
120 Constitution Drive
Menlo Park, CA 94025
Tel. (650) 812-3400

Attorneys for Plaintiff
Promptu Systems Corporation

EXHIBIT A

(12) **United States Patent**
Calderone et al.

(10) **Patent No.: US 7,260,538 B2**
(45) **Date of Patent: Aug. 21, 2007**

(54) **METHOD AND APPARATUS FOR VOICE CONTROL OF A TELEVISION CONTROL DEVICE**

(75) Inventors: **Theodore Calderone**, San Carlos, CA (US); **Mark J. Foster**, Palo Alto, CA (US); **Harry William Printz**, San Francisco, CA (US); **James Jay Kistler**, San Jose, CA (US)

(73) Assignee: **Promptu Systems Corporation**, Menlo Park, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 719 days.

(21) Appl. No.: **10/338,591**

(22) Filed: **Jan. 7, 2003**

(65) **Prior Publication Data**

US 2003/0167171 A1 Sep. 4, 2003

Related U.S. Application Data

(60) Provisional application No. 60/346,899, filed on Jan. 8, 2002.

(51) **Int. Cl.**
G10L 15/00 (2006.01)

(52) **U.S. Cl.** **704/275**; 348/14.05

(58) **Field of Classification Search** 704/275; 348/14.05

See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

5,267,323	A *	11/1993	Kimura	381/110
5,774,859	A *	6/1998	Houser et al.	704/275
5,832,439	A	11/1998	Cox	
6,253,174	B1 *	6/2001	Ishii et al.	704/231
6,397,186	B1 *	5/2002	Bush et al.	704/274
6,415,257	B1	7/2002	Junqua	
7,006,974	B2 *	2/2006	Burchard et al.	704/275

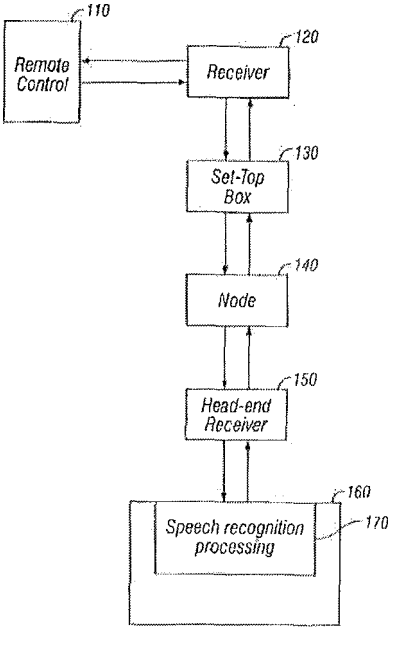
* cited by examiner

Primary Examiner—Daniel Abebe
(74) *Attorney, Agent, or Firm*—Michael A. Glenn; Glenn Patent Group

(57) **ABSTRACT**

A method and apparatus is disclosed for remotely processing voice commands for controlling a television. A voice command is uttered by a user into a microphone contained in a remote control. The voice command is digitized, modulated, compressed, and wirelessly transmitted to a wireless receiver connected to a set-top box. The voice command is then transmitted to a cable head-end unit for voice and word recognition processing. Once the command function is determined, the function is transmitted back to the set-top box where the set-top box performs the command. The microphone is activated and deactivated by pressing and releasing a push-to-talk (PTT) switch. The PTT activates other functions by being turned, double-clicked and toggled up and down, left and right.

41 Claims, 6 Drawing Sheets



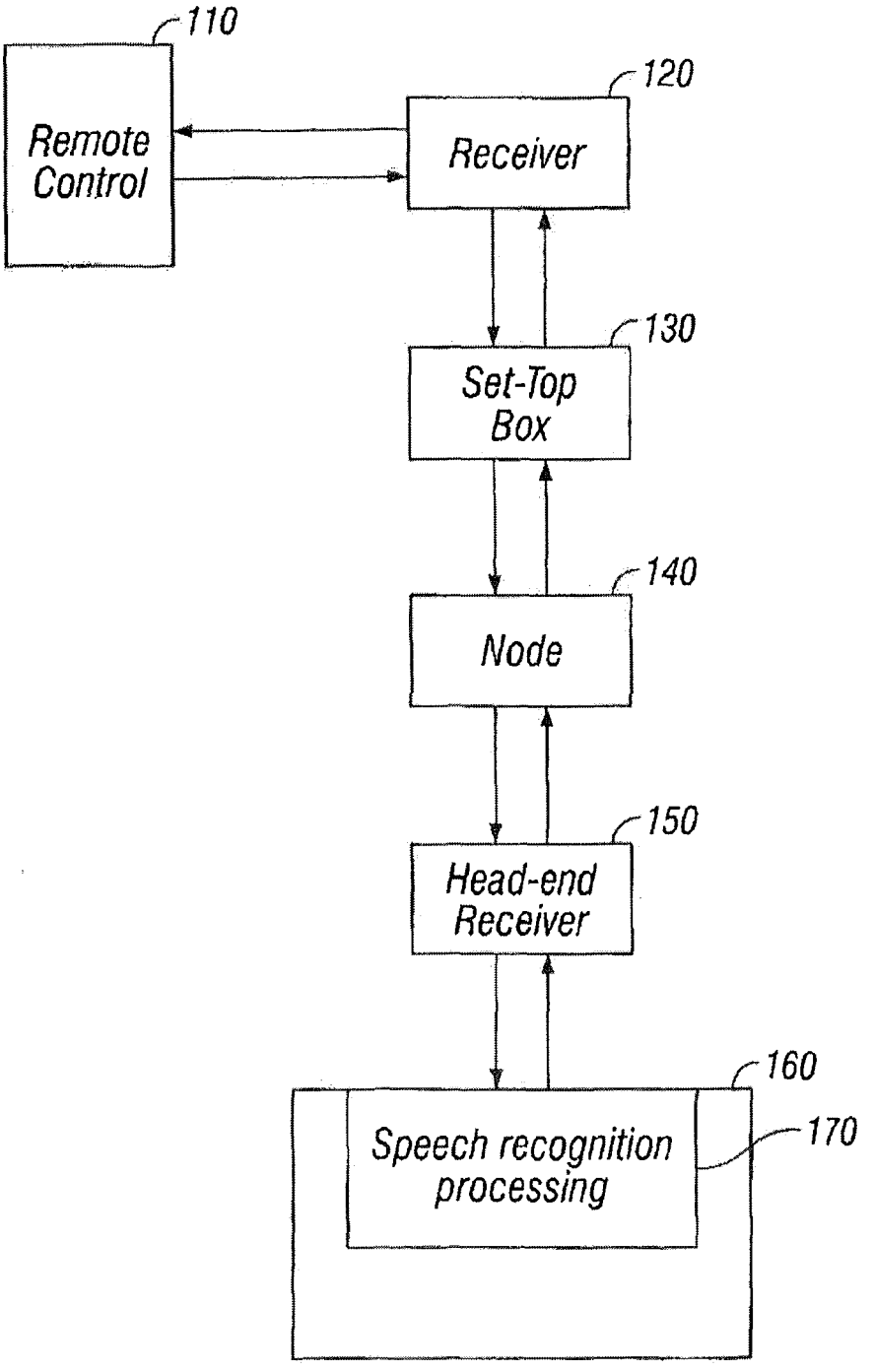

110
Remote Control
120
Receiver
130
Set-Top Box
140
Node
150
Head-end Receiver
160
Speech recognition processing
170

FIG. 1

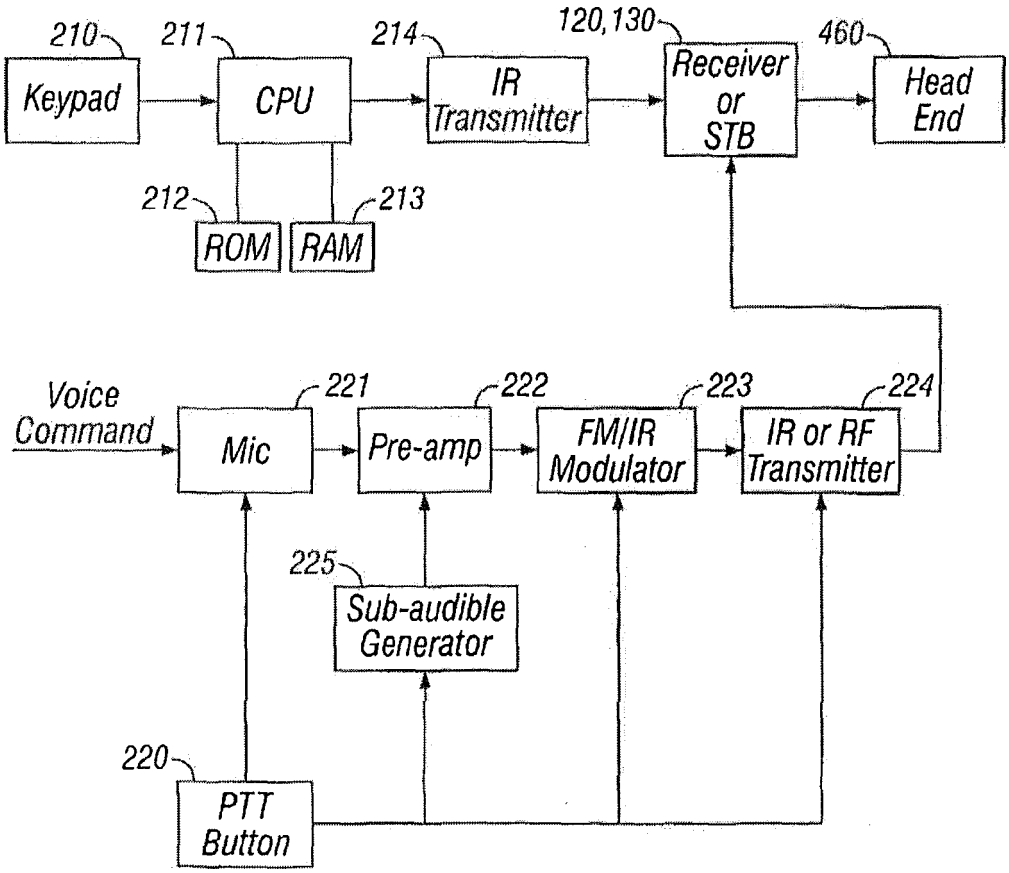

210
Keypad
211
CPU
214
IR Transmitter
120,130
Receiver or STB
460
Head End
212
ROM
213
RAM
Voice Command
221
Mic
222
Pre-amp
223
FM/IR Modulator
224
IR or RF Transmitter
225
Sub-audible Generator
220
PTT Button

FIG. 2

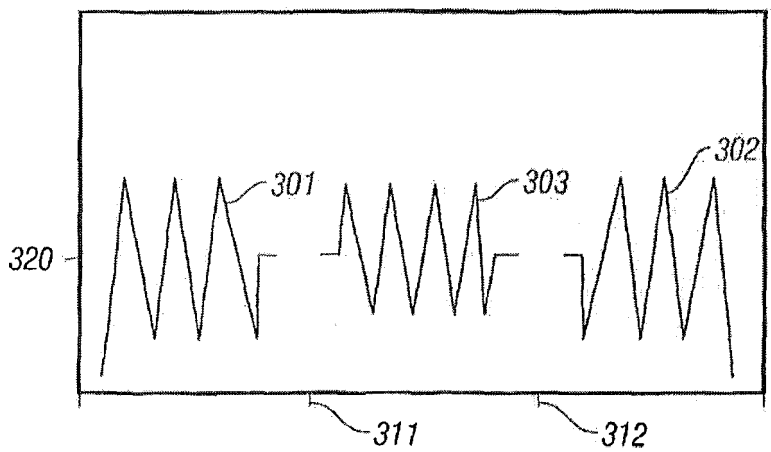

320
301
303
302
311
312

FIG. 3

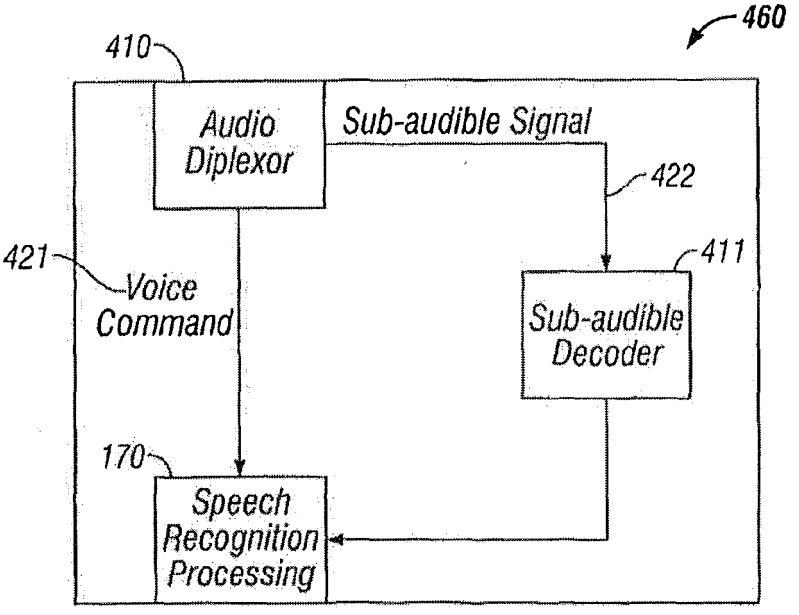

460
410
Audio Diplexor
Sub-audible Signal
422
421
Voice Command
411
Sub-audible Decoder
170
Speech Recognition Processing

FIG. 4

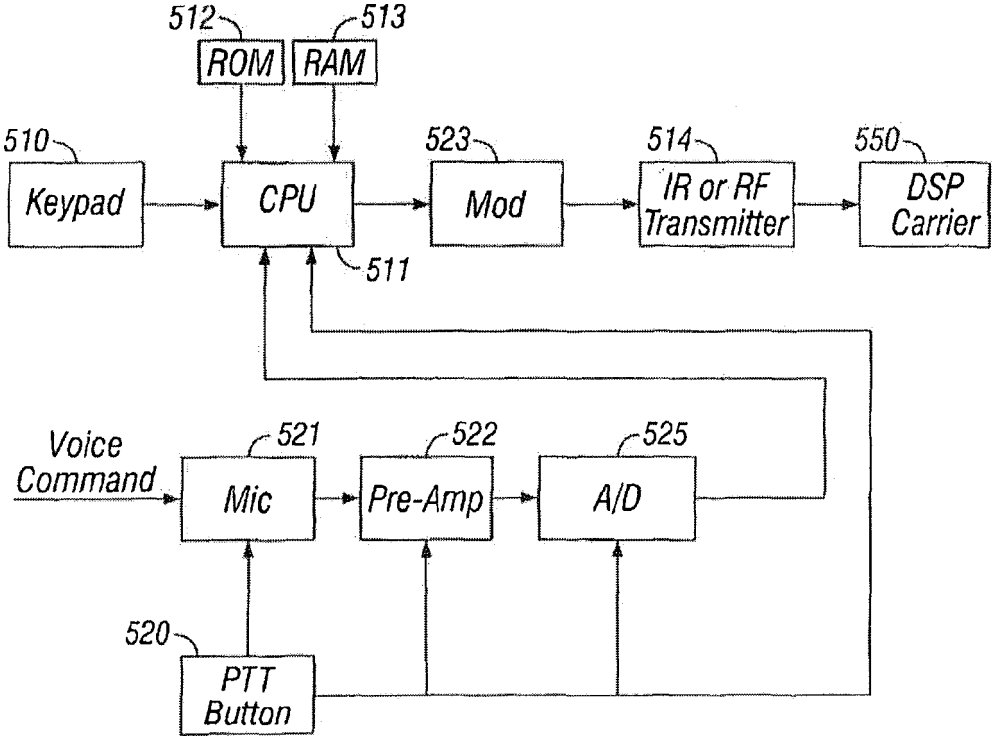

512
ROM
513
RAM
510
Keypad
CPU
523
Mod
514
IR or RF Transmitter
550
DSP Carrier
511
Voice Command
521
Mic
522
Pre-Amp
525
A/D
520
PTT Button

FIG. 5

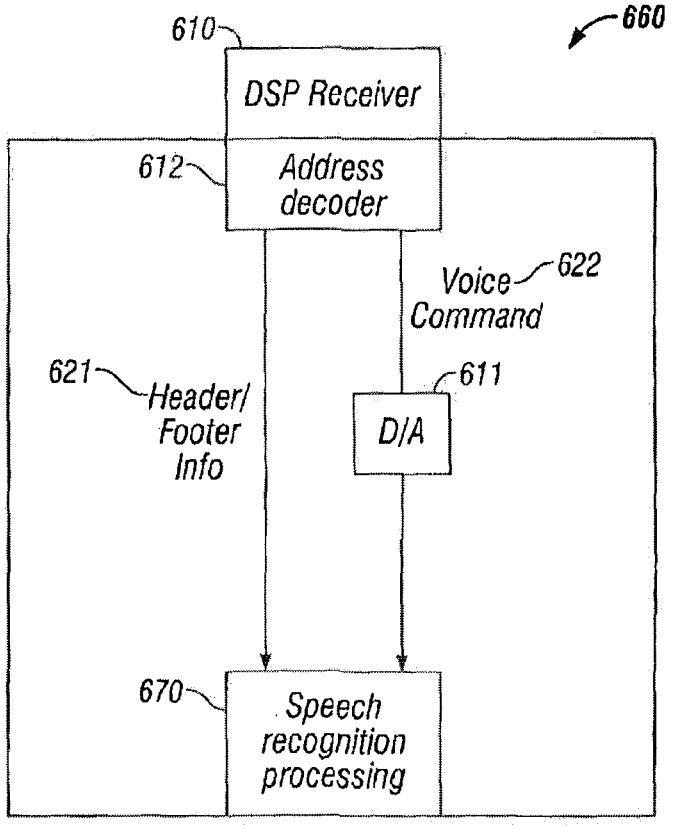

660
610
DSP Receiver
612
Address decoder
Voice Command
622
621
Header/ Footer Info
611
D/A
670
Speech recognition processing

FIG. 6

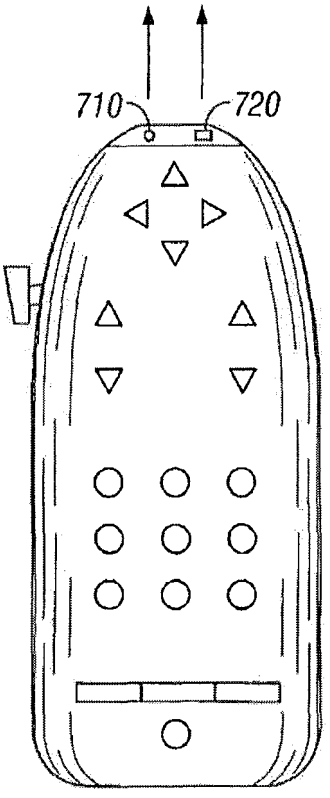


FIG. 7

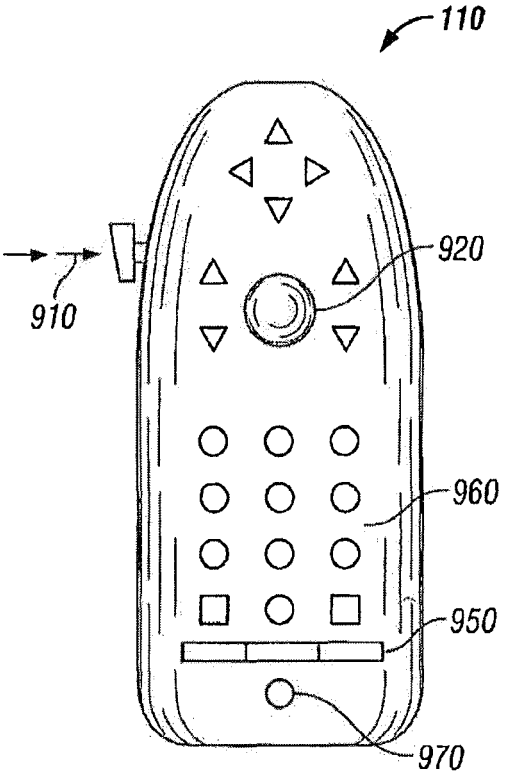


FIG. 9

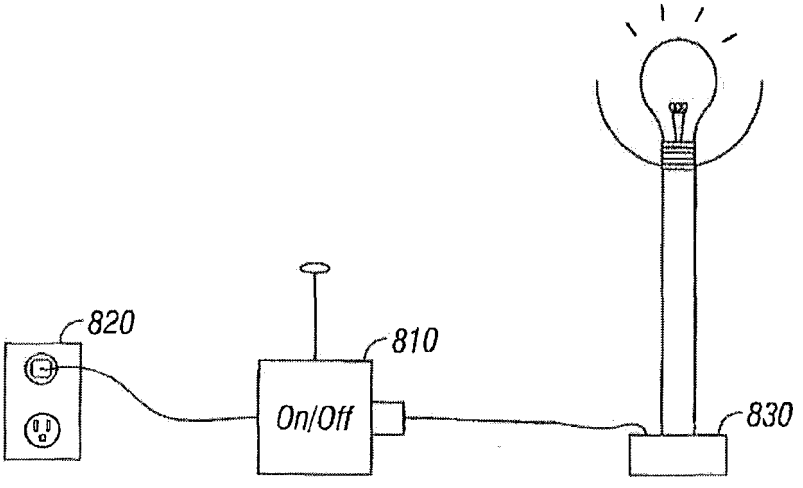


FIG. 8

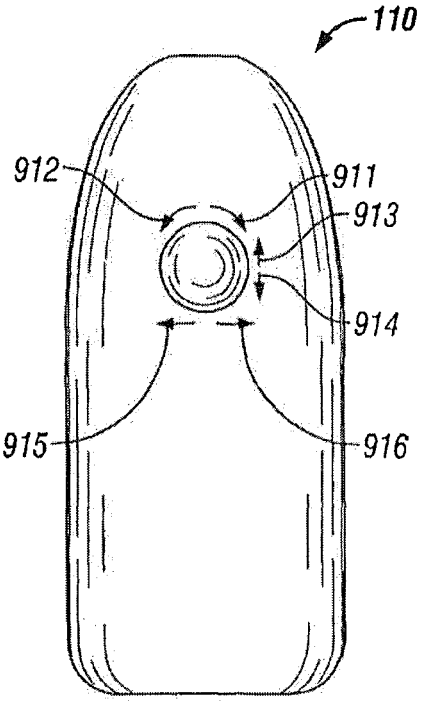

110
912
911
913
914
915
916

FIG. 10

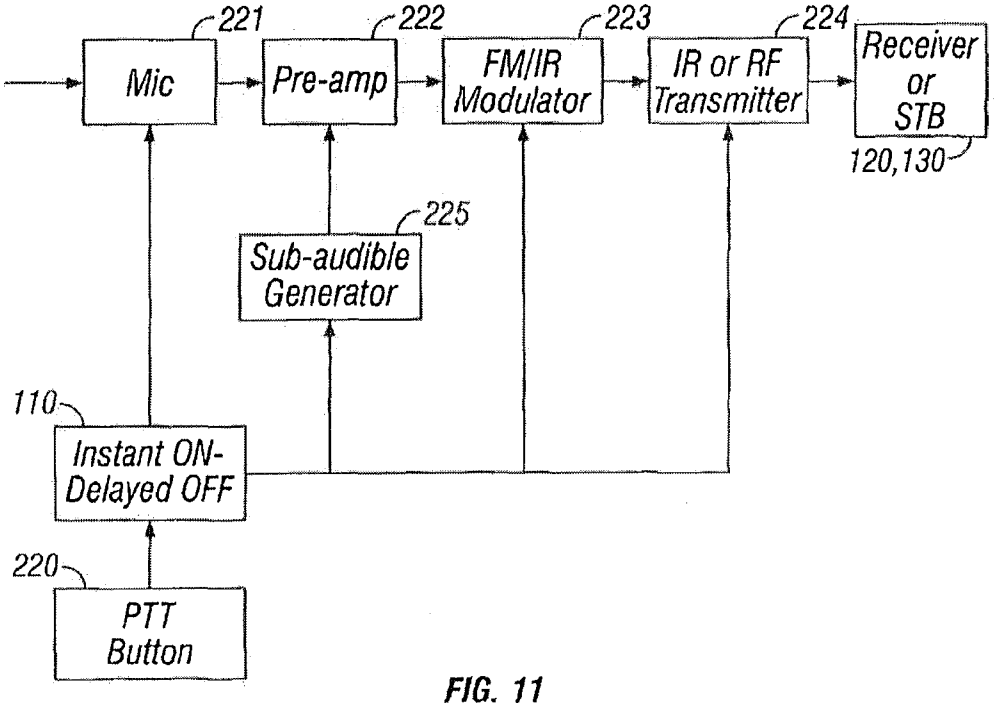

221
222
223
224
Mic
Pre-amp
FM/IR Modulator
IR or RF Transmitter
Receiver or STB
120,130
225
Sub-audible Generator
110
Instant ON-Delayed OFF
220
PTT Button

FIG. 11

METHOD AND APPARATUS FOR VOICE CONTROL OF A TELEVISION CONTROL DEVICE

CROSS-REFERENCE TO RELATED APPLICATION

This application claims priority to U.S. Provisional Patent Application No. 60/346,899 filed on Jan. 8, 2002.

BACKGROUND OF THE INVENTION

1. Technical Field

The invention relates generally to voice controlled electronic devices. More particularly, the invention relates to a voice-controlled set-top box and remote control for performing television related processes.

2. Description of the Prior Art

Speech recognition systems are capable of recognizing words spoken by human beings. Isolated word recognition systems have been developed to recognize and discriminate isolated words, i.e. words separated by a pause, which have been trained into the systems. Such s speech recognition device is summarized in D. Raj Reddy, *Speech Recognition by Machine: A Review*, Proceedings of the IEEE, April 1976, pages 501-531. Connected word recognition systems are capable of recognizing and discriminating individual words from spoken phrases.

It is also known to use voice recognition systems to control system variables of various active systems. Speech recognition control system and method, U.S. Pat. No. 4,605,080, Lemelson, Aug. 12, 1986, discloses a weighing scale involving calculating functions that include an automatic control system that can accept spoken words as input. A speech recognition system is used for processing and analyzing speech signals output by a microphone. The microphone is connected to a speech recognition computer that outputs and applies select command control signals to effect desired control functions. It is disclosed to control conveyors to place labels on containers automatically using the speech recognition system. The apparatus may be controlled to stop and start, slow down or speed up by an operator speaking the appropriate voice commands into a microphone.

Voice controlled welding system, U.S. Pat. No. 4,641,292, Tunnell , et al., Feb. 3, 1987 discloses an apparatus and method for permitting human voice control of a welding system. It is disclosed that a human operator is provided with an audio transmitter by which the operator adjusts the welding power supply through voice commands. The voice commands are issued by the operator and transmitted through a receiver to a voice recognition unit. The received signal is interpreted by a computer that is electrically connected to deliver power control signals to the welding power supply, and thereby adjusts the power delivered to the welding head. The operator may also issue voice commands to start and stop an internal combustion engine that drives the welding power supply, when such an engine is used.

A problem with the prior art voice recognition systems is that they require a sophisticated voice recognition system in close proximity to the user, requiring individual units which is quite costly. What is needed is a centralized voice command processing system such services a multitude of users.

A problem with television remote controls is their ever-increasing size. As remote controls become more sophisticated, more and more functions are added. As more functions are added, more buttons are added, causing the aforementioned increase in size. Remote controls are now approaching a practical limit. Moreover, remote control manufacturers have standard template designs for their remote controls. The standard designs dictate where the buttons are located, as well what functions will exist. What is needed is a way to add functions to an existing remote control without greatly increasing its size and without interfering with existing manufacturer designs.

A problem with voice command remote control systems is that they are activated when a sound input reaches a predetermined amplitude. Often, ambient noise reaches the pre-determined level and the system is unintentionally activated. This leads to inadvertent input and a misuse of processing power because the speech recognition unit attempts to process the noise. What is needed is way to activate and deactivate a voice command system that is not substantially affected by ambient noise.

SUMMARY OF THE INVENTION

A method and apparatus is disclosed for remotely processing voice commands for controlling a television. A voice command is uttered by a user and is received by a microphone contained in a television set-top box remote control. The voice command is modulated and wirelessly transmitted to a wireless receiver connected to the set-top box. Either before or after this transmission, it is digitized and compressed. Examples of compression algorithms used are low bit-rate encoding and conversion into cepstrals. The voice command is then transmitted, for example, to a central processing station located at a cable television head-end unit. A computer at the cable head-end unit processes the voice command for voice command recognition. Once the voice command is determined a command function is created. The command function is transmitted back to the set-top box where the set-top box performs the command function. Alternatively, the set-top box just passes on the command and the head end performs or carries out the command.

The microphone is activated by the depression of a push-to-talk (PTT) button or by word activation. Releasing the PTT button deactivates the microphone. Optionally, the PTT button can be turned to adjust the gain control on the microphone and can be toggled to activate wireless phone functions of the remote control. Double pressing the PTT button activates other functions, for example a help menu for television viewing.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a diagram illustrating elements of the voice control television system according to the invention;

FIG. 2 is a diagram illustrating a remote control for transmitting an analog voice command according to the invention;

FIG. 3 is a graph illustrating a voice command and noise;

FIG. 4 is a diagram illustrating a head-end unit for processing an analog voice command according to the invention;

FIG. 5 is a diagram illustrating a remote control for transmitting a digital voice command according to the invention;

FIG. 6 is a diagram illustrating a head-end unit for processing a digital voice command according to the invention;

FIG. 7 is a diagram illustrating a remote control that transmits a voice command through a keypad transmitter and a dedicated voice transmitter according to the invention;

FIG. 8 is a diagram illustrating a control receiver according to the invention;

FIG. 9 is a diagram illustrating a frontal view of a remote control with a multi-function push-to-talk button according to the invention;

FIG. 10 is a diagram illustrating a side view of a remote control with a multi-function push-to-talk button according to the invention; and

FIG. 11 is a diagram illustrating a remote control with an instant-on delay-off circuit according to the invention.

DESCRIPTION OF THE PREFERRED EMBODIMENT

FIG. 1 is an operational diagram illustrating a preferred embodiment of the voice control system which, for purposes of example, is an interactive television voice control system. Those skilled in the art will appreciate that other systems may be used in connection with the invention herein, such as information browsing system and representational process control systems.

A remote control **110** receives voice commands from a user through a microphone preferably located within the remote control **110**. The remote control wirelessly transmits the voice commands to a wireless receiver (RECEIVER) **120** that converts the wireless signal to a wired signal. The wireless transmission is typically infrared. Ultrasonic, radio, or other wireless transmissions are also contemplated.

The RECEIVER **120** transmits the voice command to a television controller such as a set-top box **130**. Transmissions between the RECEIVER and the set-top box may be unidirectional or bi-directional. The RECEIVER **120** transmits to the set-top box **130** when the RECEIVER **120** receives a transmission from the remote control **110**. Alternatively, the RECEIVER transmits to the set-top **130** after a specified period of time has elapsed when the RECEIVER **120** is not receiving a transmission from the remote control **110**. The set-top box **130**, RECEIVER **120** and remote control **110** may be capable of both receiving and transmitting data simultaneously. The functions of the RECEIVER **120** may also be incorporated into the television set-top box **130**.

In a preferred embodiment, the RECEIVER **120** is connected and transmits to the set-top box **130** through a serial, USB, or other connection. In another preferred embodiment, the RECEIVER **120** communicates to the set-top box **130** through the modem interface using a modem emulator and modem protocols. Set-top boxes, such as the DCT-2000 produced by Motorola of Schaumburg, Ill., provide such a modem interface. In another embodiment, the RECEIVER **120** communicates wirelessly to the set-top box **130**. A wireless receiving interface receives wireless transmission from the RECEIVER **120** and transmits the transmissions to the set-top box **130**. This may occur, for example, by means of a physical connection.

The RECEIVER **120** contains a buffer that stores the voice commands. Alternatively, the buffer is a separate unit from the RECEIVER **120**, or is contained in the set-top box **130** or the remote control **110**. Voice transmissions are either analog or digital, and are typically modulated. Modulation types include amplitude modulation, frequency modulation, pulse-position modulation (PPM), and pulse code modulation (PCM). Modulation is performed at any of the remote control **110**, the set-top box **130**, or the RECEIVER **120**. A compression scheme may be used to compress the voice command. The compression is performed at any of the remote control **110**, the set-top box **130**, the RECEIVER **120** or at a head-end unit **160**. It is also contemplated that a dedicated component may be provided that performs analog-to-digital conversions.

The voice commands are transmitted from the set-top box **130** to a node **140**, a head-end receiver **150** and to a head end unit **160** along a cable television uplink, where speech recognition processing **170** is performed. In the preferred embodiment, speech recognition processing includes user voice identification and word recognition. Word recognition includes probability or semantic analysis checking especially where a voice command contains multiple words. The probability checking checks that the words logically go together.

The central processing station is designed to handle a multitude of voice command inputs from a multitude of cable television users. After the voice command is processed, the central processing station **160** sends a corresponding command function to the cable set-top box **130** or other system component where the command is then performed.

FIG. 2 illustrates one embodiment of the remote control, where an analog signal is transmitted from the remote control to a head end unit **460** for processing. The remote control **110** may include a keypad **210** that allows a user to perform such tasks as entering numerical values, adjusting the volume, and changing the channels. The keypad entries are processed by a CPU **211** that has a ROM **212** and RAM **213** electronically coupled to it. The CPU functions are typically either hardware implemented or an embedded software solution. The processed keypad entries are transmitted by infrared **214** to a RECEIVER **120**, or to a set-top box **130** that can receive infrared transmissions.

A push-to-talk (PTT) button **220** activates the microphone **221** on the remote control, as well as activating a pre-amp **222**, modulator **223**, transmitter **224**, and optionally, a sub-audible generator **225**. The PTT button **220** brackets the voice command by indicating a beginning when depressed, and an end when released. The bracketing of the speech provides a higher signal to noise ratio because unintended noise at the beginning and end of the transmissions are eliminated. When the PTT button **220** is released, the microphone **221**, pre-amp **222**, modulator **223**, transmitter **224** and sub-audible generator **225** are deactivated optionally after a suitable delay.

A voice command received by the microphone **221**, is processed by the pre-amp **222**, modulated **223**, and transmitted **224** to the receiver **120** or set-top box **130**. If the transmitter **224** is a radio frequency (RF) transmitter, then the modulator **223** is, for example, an FM modulator. If the transmitter **224** is an infrared transmitter, then the modulator **223** is an infrared modulator.

If present, the sub-audible generator **225** produces a sub-audible tone. The voice command and sub-audible signal are combined in the pre-amp **222**. The voice command and sub-audible signal may also be combined in either in the set-top box **130** or in the RECEIVER **120**. The sub-audible tone indicates to the receiver or head-end unit **460** that the microphone is on air and that the user is producing voice commands. Alternatively, the on-air status of the microphone may be detected by other means, such as carrier detection. The combined signal is transmitted to the RECEIVER **120** or to a set-top box **130** that can receive wireless transmissions.

FIG. **3** is graph that illustrates a voice command **303** and noise **301**, **302** that precedes and comes after the voice command. In a typical voice activated system, the microphone is activated when an audible sound reaches predetermined amplitude **320**. Thus, in an amplitude activated system, the noise **301**, **302** and voice command **303** all activate the microphone. However, with a PTT button, only the depression **311** of the button activates the microphone, and the release **312** of the button deactivates the microphone. Thus the microphone does not receive the unwanted noise **301**, **302**.

FIG. **4** illustrates a head-end unit **460** for processing an analog voice command. An audio diplexor **410** receives the transmissions and divides the signal into the voice command **421** and the sub-audible signal **422**. A sub-audible decoder **411** processes the sub-audible signal **422**. The sub-audible signal **422** contains information, such as the source of the transmission, and also indicates that the transmission is active. The head-end unit processes **170** the voice command and decoded sub-audible signal information, and determines a command function

FIG. **5** illustrates a remote control that transmits a digital signal to a head-end unit **660** for processing. Keypad entries **510** are processed by a CPU **511** to which a ROM **512** and RAM **513** is electronically coupled. The processed keypad entries are modulated **523** and transmitted **514** by infrared or RF to a digital server provider (DSP) carrier **550**.

A PTT button **520** activates a microphone **521**, pre-amp **522**, and analog-to-digital converter **525**. The microphone **521** receives a voice command and transmits it to a pre-amp **522**. The pre-amp **522** transmits the voice command to the analog-to-digital converter **525** where the voice command is converted from analog to digital. The digital voice command is transmitted to the CPU **511** where header and footer information is added. A unique digital signature in the header identifies the remote control. Header information is typically stored in the ROM **512**. The CPU **511** transmits the digital command to a digital service provider (DSP) carrier **550** either through an infrared or RF transmitter. The DSP carrier **550** is part of either the wireless receiver, the set-top box or may be a separate unit.

In considering FIG. **2** and FIG. **5**, it is noted that a hybrid system is also envisioned, in which the remote control transmits analog speech to the receiver where it is digitized and sent in digital format to the head-end.

FIG. **6** illustrates a head-end unit **660** for processing a digital voice command. A digital DSP receiver **610** receives the digital signal and transmits the signal to an address decoder **612**. The address decoder separates and interprets the header and footer information **621** from the digital voice command **622**. The voice command portion **622** of the signal is transmitted to a digital-to-analog converter **611**. The digital-to-analog converter **611** transmits the analog voice command to a speech recognition processor **670** where it analyzed for voice command recognition, and a command function is derived.

A transmission is cepstral analyzed for the purposes of facilitating the computer analysis of speech for voice recognition. Cepstral analysis is a method of feature extraction that is known in the art. Cepstral analysis, or feature extraction, is performed on a digitized speech signal and results in a representation of the signal that characterizes the relevant features of the speech. It can be regarded as a data reduction procedure that retains vital characteristics of the speech and eliminates undesirable interference from irrelevant characteristics of the digitized signal, thus easing the decision-making process of a computer. An example of a preferred method of cepstral analysis is disclosed in Wang, Method of training neural networks used for speech recognition, U.S. Pat. No. 5,509,103, Apr. 16, 1996.

Typically, a remote control contains an existing infrared transmission method for transmitting keypad entries to a set-top box. Using the existing infrared transmitter alone is insufficient for voice commands as the transmission rate is approximately 1200 baud. Thus, a dedicated transmission method is used to transmit the voice commands from the remote control to the receiver.

Examples of transmission methods include infrared, FM radio, AM radio and ultrasonic. In one embodiment, a high digital signal rate is used for the transmissions, such as burst transmissions. In another embodiment, a high data rate is achieved by sending multiple signals simultaneously in the form of a multiplexed signal.

In a preferred embodiment, Bluetooth transceivers are used to facilitate communication between the remote control **110** and the RECEIVER **120**. Bluetooth defines a universal radio interface in the 2.45 GHz Industrial-Scientific-Medical (ISM) frequency band. Specific portions of this bandwidth are available on an essentially global basis. Thus, Bluetooth-capable systems operate internationally. Bluetooth permits disparate electronic devices or systems to communicate with each other via short-range communications. A group of two or more devices in local communications with each other using Bluetooth, form a Bluetooth network, referred to as a piconet. A piconet comprises up to eight Bluetooth devices, with one device serving as the master and the remaining devices acting as slaves in the piconet. A given Bluetooth device in a given piconet may alternately participate in other piconets, with a group of piconets referred to as a scatternet.

While those skilled in the art may understand and practice the present invention absent detailed presentation of Bluetooth specifications, the papers *Bluetooth—The universal radio interface for ad hoc, wireless connectivity*, by Jaap Haartsen, Ericsson Review No. 3, 1998. pp. 110-117, and *Ericsson's Bluetooth Modules*, by Arfwedson and Sneddon, Ericsson Review No. 4, 1999. pp. 198-205, provide substantial detail regarding the implementation and operation of Bluetooth transceivers. Additional Bluetooth technical specifications may be obtained by accessing a Bluetooth Special-Interest-Group (SIG) website at http://www.bluetooth.com.

FIG. **7** illustrates another preferred embodiment where a voice command is simultaneously transmitted over both a dedicated voice transmitter **710** and an existing keypad transmitter **720**. Transmitting with both the dedicated and existing transmitter increases the baud rate of the transmission. In one embodiment, the voice command transmission is divided between the two transmitters.

In another embodiment, where the remote control transmits an analog signal, the existing keypad transmitter **720** transmits the sub-audible PTT signal, while the dedicated voice transmitter **710** transmits the voice command.

In another embodiment, the transmitters transmit at different frequencies. The RECEIVER has two receivers. One receiver receives transmissions at the dedicated voice transmitter frequency, while the other receiver receives transmissions at the existing keypad transmitter frequency.

In yet another embodiment the transmitters transmit at the same frequency. Each transmission is out of phase with the other. A single receiver at the RECEIVER receives both signals. Each signal has a header that defines how the transmissions are constructed, and how they should be combined.

The remote control **110** may have the ability for bi-directional communication with the RECEIVER **120**. If so, the bi-directional communication allows the transmission of information from the RECEIVER **120** to the remote control **110**. The bi-directional communication may be accomplished by using the Bluetooth standard when radio transceivers are used, and by TWIRP, produced by SolutioNet, Ltd. of Williston, Vt., when infrared receivers and transmitters are used. Transmissions sent to the remote control are typically command functions. The command functions are used to control remote control functions such as the gain control.

The command functions may also be used to control the functions of other electrical devices. In one embodiment, the remote control relays command functions to electrical devices that are already capable of being controlled by a wireless control, for example, a stereo. In this example, the command functions include turning the stereo on and off, adjusting the volume, switching stereo functions and changing radio stations. Alternatively, the command functions may be relayed via a wired "IR blaster" transmitter connected to theremote or set-top box.

FIG. **8** illustrates another embodiment where the command functions control electrical devices that do not have an existing wireless control method. The remote control transmits command functions to a control receiver **810** that is connected in-line between an electrical device such as a light **830** and a power outlet **820**. The command functions turn the power on and off to the device. For example, a user wishes to turn a light on. The user would say into the remote control, "lights on." The head-end unit would process the voice command and send a control function back to the remote control or IR blaster. The remote control transmits the control function to the control receiver **810** where it switches the power on to the light **830**.

In another preferred embodiment a word recognition unit activates the microphone upon the speaking of a specific word. For example, when the word "agile" is spoken, the microphone is turned on. Once turned on, the microphone stays on continuously during speech, and turns off five seconds after the last word is spoken. The word recognition unit is typically constructed of a voice processor and a buffer. The unit is activated when the amplitude of speech reaches a predetermined level, similar to voice activation devices found in tape recorders. In this way, the remote control uses less power because the word recognition unit is not always on. Also, to reduce costs and weight, a low level processor is used with limited speech recognition. However, as processor speeds increase, and die sizes are cast smaller, more powerful processors may be used.

In another embodiment of the invention, low-bit rate audio encoding is used for non-repudiation purposes. Non-repudiation provides a method to guarantee that a party to a transaction cannot falsely claim that they did not participate in a transaction. Traditionally, handwritten signatures are used to ensure this. When a consumer writes a check, presenting a driver's license ensures the identity of the writer, i.e. authentication, while the signature on the check ensures that the consumer was in fact present and agreed to write the check, i.e. non-repudiation. When a user utters a voice command, the low-bit rate encoding is used to mark the transaction. Such marking is typically used when a user orders a pay-per-view event or other transaction involving goods or services for a fee. The low bit-rate recording serves as a non-memory intensive way of recording the user's voice command.

Non-repudiation is also performed by identifying the source of the voice command through an identification code that is attached in the header of a voice command. A user may also be required to enter on the remote control keypad an identifying code, such as the user's home phone number or "PIN" number.

FIGS. **9** and **10** illustrate another preferred embodiment of the invention, where multiple functions are associated with the PTT button **910**. The primary function of the PTT button is to engage and disengage a command pathway for a voice command from the user to the head-end unit. Other functions are also activated by manipulation of the PTT button.

An on-screen help menu is activated by double clicking the PTT button **910** in rapid succession to provide, for example, a help screen. The help screen typically provides information on how to use the system and also provides program information.

Optionally, the PTT button **910** may also adjust the sensitivity of the microphone. A clockwise turn **911** of the button increases the sensitivity while a counter-clockwise turn **912** decreases it. The sensitivity can be decreased to a degree such that the microphone is muted. An indicator **950** is located on the remote control or receiver can be used to indicate the amplitude of a user's voice command. The indicator lets a user know whether his voice command meets the minimum amplitude required for an accurate and process able recording. The PTT button **910** also functions as a shift button. Thus, depressing the PTT button **910** along with pressing keypad buttons **960** on the remote control, gives the keypad buttons **960** a secondary function.

Optionally, the remote control may also functions as a wireless phone. Toggling the PTT button upward **913** turns the phone on, and a user is able to dial a phone number using the remote control keypad. Toggling the button downward **914** turns off the phone. After a user releases the PTT button from its toggled position, the PTT button returns to a middle, default position. The microphone **970** functions as a phone receiver and a speaker **920** on the remote control allows the user to hear a phone transmission. The RECEIVER functions **120** as the base station for wireless phone functions. The RECEIVER **120** connects to a phone jack using a standard telephone wire. The wireless phone transmissions are transmitted to the receiver using the same communication pathways that the voice commands use.

In one embodiment of the invention, infrared repeaters are used to improve the performance of an infrared remote control. Infrared repeaters are placed throughout the operating environment of the remote control and RECEIVER. When a user uses a keypad function, it is natural for a user to point the remote control at the RECEIVER of the set-top box making the need for infrared repeaters unnecessary. However, when speaking into the remote control, a user's tendency is to place his mouth in close proximity to the microphone on the remote control. In most cases the infrared transmitter is not aimed towards the RECEIVER. The placement of the infrared repeaters ensures that the RECEIVER receives the infrared transmission from the remote control.

In a preferred embodiment the PTT button is a metal switch that activates the microphone transmission functions by engaging mechanical contacts. Switches of this type suffer from bounce. Hardware and software solutions exist for debouncing the switch. One example of a hardware debouncing solution is a circuit that has an RC time constant that swamps out the bounce. A typical software solution runs a routine that kills time long enough to allow the contacts to stop bouncing. These examples are not meant to be exhaustive but merely illustrative of what exists in the art.

Often, a user releases the PTT button prior to finishing a voice command. FIG. **11** illustrates a timing circuit **1110** with an analog remote control transmitter. The timing circuit keeps the microphone **221**, sub-audible generator **225**, pre-amp **222**, modulator **223**, and transmitter **224** activated for a period of time after the user has released the PTT button **220**. The period of time is typically from one hundred milliseconds to one second. The timing circuit is activated only when the PTT button is depressed for more than some minimal time. This enables fast clicks of the PTT button to control other functions.

Although the invention is described herein with reference to the preferred embodiment, one skilled in the art will readily appreciate that other techniques and applications may be substituted for those set forth herein without departing from the spirit and scope of the present invention. Accordingly, the invention should only be limited by the claims included below.

The invention claimed is:

1. A method for providing voice recognition processing at a cable television head-end unit for a plurality of voice controlled television cable set-top boxes in a cable television network, comprising the steps of:

a television remote control receiving user-activated indication of a voice command;

receiving said voice command through a microphone associated with said television remote control;

said television remote control wirelessly transmitting a signal representing said voice command to a cable set-top box;

said cable set-top box transmitting a signal representing said voice command via cable television link to a remotely located head-end unit;

processing said voice command at said head-end unit;

the head-end unit deriving a set-top-box-compatible command function corresponding to said voice command;

the head-end unit transmitting said command function to said cable set-top box via the cable television link;

performing said command function at said cable set-top box.

2. A method for providing voice recognition processing at a cable television head-end unit for a plurality of television controllers in a cable television network, comprising the steps of:

a television remote control receiving indication of a voice command by user operation of an activation feature associated with the television remote control;

receiving said voice command through a microphone associated with said television remote control;

said television remote control wirelessly transmitting a signal representing said voice command to a television controller;

said television controller transmitting a signal representing said voice command via cable television link to a remotely located cable head-end unit;

the head-end unit deriving a controller-compatible command function corresponding to said voice command;

the head-end unit transmitting said command function to said controller via the cable television link; and

the controller performing said command function.

3. The method of claim **2**, wherein the activation feature comprises a push-to-talk switch and activation of said push-to-talk switch comprises any of:

pressing said push-to-talk switch, double pressing said push-to-talk switch, turning said push-to-talk switch clockwise and/or counterclockwise, and toggling said push-to-talk switch.

4. The method of claim **2**,

where the activation feature comprises a push-to-talk switch, and the steps further comprise, responsive to operation of the push-to-talk switch, performing functions including any of:

activating said microphone, deactivating said microphone, activating a help menu, deactivating a help menu, activating a telephone, deactivating said telephone, adjusting the gain on said microphone, muting said microphone and performing a shift function.

5. The method of claim **2**,

where user operation of the activation feature comprises making one or more predetermined utterances to operate a voice-activation feature of the remote control.

6. The method of claim **2**, wherein said remote control comprises a wireless transmitter employing any of the following wireless communication formats:

infrared, ultrasonic, radio frequency transmission, burst transmission, a multiplexed signal, a transmission device that conforms to a Bluetooth Specification, and an infrared repeater.

7. The method of claim **2**,

where the television remote control includes a keypad and a first transmitter to wirelessly transmit signals representing user entries upon the keypad;

where the operation of wirelessly transmitting the signal representing the voice command to the television controller employs a second transmitter separate from the first transmitter.

8. The method of claim **2**,

where the television remote control includes a keypad and a transmitter to wirelessly transmit signals representing user entries upon the keypad;

where the operation of wirelessly transmitting the signal representing the voice command to the television controller employs the transmitter;

where the steps further comprise distinguishing between signals representing the voice commands and the signals representing user entries upon the keypad by one or more of the following: using transmissions of same frequency and different phase, using different frequencies, using different header or footer information, including a sub-audible signal in one or both signals.

9. The method of claim **2**,

where the steps further comprise:

the television remote control adding information to the signal representing the voice command transmitted to the television controller, said information contained in locations comprising any of:

header data, footer data, a sub-audible signal.

10. The method of claim **9** wherein said information comprises any of:

source of said voice command, identity of a user, on air signal, off air signal, beginning of a voice command, and end of a voice command.

11. The method of claim **2**, further comprising converting signals representing said voice command into cepstrals, the converting step performed by any of:

said remote control, said television controller, a wireless transmission receiver that receives signals from said remote control.

12. The method of claim **2**, performing non-repudiation measures comprising any of:

deriving a low bit rate codec of said voice command, determining the source of said voice command, analyzing said voice command for the identity of the user, and requiring the input of a code.

13. The method of claim 2, further comprising:

automatically delaying deactivation of the activation feature for a period of time after user activation of the activation feature.

14. The method of claim 2, further comprising the step of:

the head-end unit transmitting further command functions to the television controller for relay to the television remote control;

responsive to receiving the further command functions, the television controller wirelessly forwarding the further command functions to the remote control.

15. The method of claim 14,

further comprising:

responsive to receiving the further command functions, the remote control performing any of: adjusting the gain control of said microphone and muting said microphone.

16. The method of claim 14, further comprising the step of:

responsive to receiving the further command functions, the remote control wirelessly transmitting said command function to a wirelessly operated device.

17. The method of claim 2, where the step of deriving the controller-compatible command function includes any of: word recognition, semantic analysis.

18. An apparatus for providing voice recognition processing at a cable television head-end unit for a plurality of voice controlled television cable set-top boxes in a cable television network, comprising:

a television remote control including: activation means for receiving user-activated indication of a voice command, microphone means for receiving the voice command, and transmission means for wirelessly transmitting a signal representing the voice command to a cable television controller;

a cable television controller including receiver means for receiving the signal representing the voice command from the television remote control and transmitter means for transmitting a signal representing the voice command via cable television link to a remotely located head-end unit;

a head-end unit including processing means for deriving cable-television-controller-compatible command functions corresponding to signals representing the voice commands received from the cable television controllers, and transmission means for transmitting signals representing the command functions back to respective cable television controllers;

where the cable television controller additionally includes second receiver means for receiving the signals representing the command functions from the head-end unit via the cable television link, and where the cable television controller includes means responsive to receipt of the command functions for executing the command functions.

19. An apparatus for providing voice recognition processing at a cable television head-end unit for a plurality of television controllers in a cable television network, comprising:

a television remote control including: activation feature to receive user indication of a voice command, a microphone to receive the voice command, and a transmitter to wirelessly transmit a signal representing the voice command to a cable set-top box;

a cable set-too box including a receiver to receive the signal representing the voice command from the television remote control and a transmitter to transmit a signal representing the voice command via cable television link to a remotely located head-end unit;

a head-end unit to derive set-top-box-compatible command functions corresponding to signals representing voice commands received from the cable set-top-boxes and transmit signals representing the command functions back to respective cable set-too boxes;

where the cable set-top box additionally includes a second receiver to receive signals representing the command functions from the head-end unit via the cable television link, the cable set-top box responsive to receipt of the signals representing the command functions to execute the command functions.

20. The apparatus of claim 19, wherein said activation feature comprises a push-to-talk switch activated by any of:

pressing said push-to-talk switch, double pressing said push-to-talk switch, turning said push-to-talk switch clockwise and counterclockwise, and toggling said push-to-talk switch.

21. The apparatus of claim 19,

where the activation feature comprises a push-to-talk switch with functions comprising any of:

activating said microphone, deactivating said microphone, activating a help menu, deactivating a help menu, activating a telephone, deactivating said telephone, adjusting the gain on said microphone, muting said microphone, and performing a shift function.

22. The apparatus of claim 19,

where the activation feature is operated by receipt of one or more predetermined utterances.

23. The apparatus of claim 19, wherein said remote control comprises a wireless transmitter employing any of the following wireless communication formats:

infrared, ultrasonic, radio frequency transmission, burst transmission, a multiplexed signal, a transmission device that conforms to a Bluetooth Specification, and an infrared repeater.

24. The apparatus of claim 19,

where the television remote control includes a keypad and a first transmitter to wirelessly transmit signals representing user entries upon the keypad;

where the operation of wirelessly transmitting the signal representing the voice command to the cable set-top box employs a second transmitter separate from the first transmitter.

25. The apparatus of claim 19,

where the television remote control includes a keypad and a transmitter to wirelessly transmit signals representing user entries upon the keypad;

where the operation of wirelessly transmitting a signal representing the voice command to the cable set-top box employs the transmitter;

where the television remote control is configured to distinguish the signals representing the voice commands and the signals representing user entries upon the keypad by one or more of the following: using transmissions of same frequency and different phase, using different frequencies, using different header or footer information, including a sub-audible signal in one or both signals.

26. The apparatus of claim 19,

where the television remote control is configured to add information to the signal representing the voice command transmitted to the cable set-top box, where said information is contained in locations comprising any of:

header data, footer data and a sub-audible signal.

27. The apparatus of claim 26 wherein said information comprises any of:

source of said voice command, identity of a user, on air signal, off air signal, beginning of a voice command, and end of a voice command.

28. The apparatus of claim 19,

where one of the following is configured to convert signals representing the voice command into cepstrals:

said remote control, said cable set-top box, a wireless transmission receiver that receives said transmission from said remote control.

29. The apparatus of claim 19, the television remote control further configured to automatically delay deactivation of the activation feature for a period of time after user activation.

30. The apparatus of claim 19, where:

the head-end unit is configured to transmit further command functions to the set-top box for relay to the remote control;

the set-top box is configured, responsive to receiving the further command functions, to forward the further command functions to the remote control.

31. The apparatus of claim 30,

where the remote control is configured, response to receiving the further command functions, to perform any of: adjusting the gain control of said microphone and muting said microphone.

32. The apparatus of claim 30,

the remote control is further configured, responsive to receiving the further command functions, to wirelessly transmit said command functions to a wirelessly operated device.

33. The apparatus of claim 19, where the head-end unit is programmed such that deriving the set-top-box-compatible command function includes any of: word recognition, semantic analysis.

34. A centralized multi-user voice operated television control system, comprising:

television remote controls configured to directly and wirelessly control television sets and additionally to receive user voice input and wirelessly transmit first output representative of the voice input to television set-top boxes;

television set top boxes configured to receive television input signals via cable television link and provide television output signals compatible with television sets, the set top boxes additionally responsive to receiving the first output from the television remote controls to transmit representative second output to a central processing station via the cable television link;

a centralized processing station configured to receive and process second output from a multitude of television set top boxes by applying voice recognition to the second output to identify user-intended voice commands, to derive set-top-box-compatible instructions to carry out the identified voice commands, and returning signals representing the instructions to respective top boxes via the cable television link;

where the set top boxes are further responsive to receiving the signals representing the instructions from the central processing station to execute the instructions.

35. The system of claim 34, where the television remote controls include circuitry to ignore user voice input except during activation of a user-operated activation feature.

36. The system of claim 34, where each television remote control includes a signature generator providing a predetermined signature signal concurrent with wireless transmission of the first output, and where the set top boxes are configured to prevent transmission to the central processing station of output from the television remote controls not containing the predetermined signature signal.

37. The system of claim 34,

where each television remote control includes a keypad and a first transmitter to wirelessly transmit signals representing user entries upon the keypad;

where the operation of wirelessly transmitting the first output employs a second transmitter separate from the first transmitter.

38. The system of claim 34,

where each television remote control includes a keypad and a transmitter to wirelessly transmit signals representing user entries upon the keypad;

where the operation of wirelessly transmitting the first output employs the transmitter;

where each television remote control is further configured to distinguish the first output from signals representing user entries upon the keypad by one or more of the following: using transmissions of same frequency and different phase, using different frequencies, using different header or footer information, including a sub-audible signal in one or both signals.

39. The system of claim 34, where:

the set top boxes are configured to only execute certain identified commands, and for other commands, to wirelessly forward a representative fourth output representing the other commands to the television remote controls;

the television remote controls further include receivers to receive the fourth output, and are further configured to directly control wireless devices by wirelessly transmitting signals representing the fourth output.

40. A centralized multi-user voice operated television control system, comprising:

a plurality of television remote control means each for directly and wirelessly controlling television sets and additionally receiving user voice input and wirelessly transmitting first output representative of the voice input to a television set-top box means;

a plurality of television set top box means each for receiving television input signals via cable television link and providing television output signals compatible with television sets, and responsive to receiving the first output from one of the television remote control means to transmit representative second output to a central processing station via the cable television link;

a centralized processing station configured to receive and process second output from a multitude of television set top box means by applying voice recognition to the second output to identify user-intended voice commands, to derive set-top-box-means-compatible instructions to carry out the identified voice commands, and returning signals representing the instructions to respective set top box means via the cable television link;

where each set top box means is further responsive to receiving the signals representing the instructions from the central processing station to execute the instructions.

41. A method for operating a centralized multi-user voice operated television control system that includes a multitude of television remote controls situated at various television

viewing sites, a multitude of set top boxes situated at the television sites to receive television input signals via cable television link and provide television output signals compatible with television sets at the television viewing sites, and a centralized processing station remote from the television viewing sites and coupled to the set top boxes via the cable television link, the method comprising operations of:

operating the television remote controls to perform additional operations including receiving user voice input and wirelessly transmitting first output representative of the voice input to set-top boxes;

operating the set top boxes to perform additional operations including, responsive to receiving the first output from the television remote controls, transmitting representative second output to a central processing station via the cable television link;

operating the centralized processing station to receive and process second output from a multitude of set top boxes by applying voice recognition to the second outputs to identify user-intended voice commands, to derive set-top-box compatible instructions to carry out the identified voice commands, and to return signals representing the instructions to the set top boxes via the cable television link;

operating the set top boxes to perform further operations including, responsive to receiving the signals representing the instructions from the central processing station, executing the instructions.

* * * * *

EXHIBIT B

US007047196B2

(12) **United States Patent**
Calderone et al.

(10) **Patent No.: US 7,047,196 B2**
(45) **Date of Patent: May 16, 2006**

(54) **SYSTEM AND METHOD OF VOICE RECOGNITION NEAR A WIRELINE NODE OF A NETWORK SUPPORTING CABLE TELEVISION AND/OR VIDEO DELIVERY**

(75) Inventors: **Theodore Calderone**, San Carlos, CA (US); **Paul M. Cook**, Woodside, CA (US); **Mark J. Foster**, Palo Alto, CA (US)

(73) Assignee: **AgileTV Corporation**, Menlo Park, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 1138 days.

(21) Appl. No.: **09/785,375**

(22) Filed: **Feb. 16, 2001**

(65) **Prior Publication Data**

US 2001/0056350 A1 Dec. 27, 2001

Related U.S. Application Data

(63) Continuation-in-part of application No. 09/708,315, filed on Nov. 7, 2000, now Pat. No. 6,480,703, and a continuation-in-part of application No. 09/679,115, filed on Oct. 4, 2000, and a continuation-in-part of application No. 09/664,874, filed on Sep. 19, 2000, now abandoned, and a continuation-in-part of application No. 09/661,486, filed on Sep. 14, 2000.

(60) Provisional application No. 60/210,440, filed on Jun. 8, 2000.

(51) **Int. Cl.**
G10L 21/00 (2006.01)

(52) **U.S. Cl.** **704/270.1**; 704/201; 704/231; 715/719; 715/512; 715/726; 709/217

(58) **Field of Classification Search** 704/270–278, 704/231, 251, 246, 201; 715/719, 512, 513, 715/716, 721, 723, 726, 733; 709/217
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,506,387 A 3/1985 Walter 455/612

FOREIGN PATENT DOCUMENTS

EP 0 782 337 7/1997 ..... 7/173

(Continued)

OTHER PUBLICATIONS

ACTV Awarded Three Key Digital Television Patents: Coverage Extended to ACTV's Core Technology for Eliminating Video Switching Delay; May 8, 2001; PRNewswire.

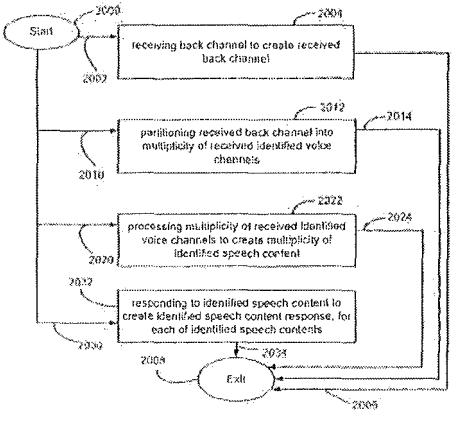
(Continued)

Primary Examiner—Vijay Chawan
(74) *Attorney, Agent, or Firm*—Michael A. Glenn; Glenn Patent Group

(57) **ABSTRACT**

A method and system of speech recognition presented by a back channel from multiple user sites within a network supporting cable television and/or video delivery is disclosed. The preferred embodiment of the invention comprises ans system and method of using a back channel containing a multiplicity of identified speech channels from a multiplicity of user sites presented to a speech processing system at a wireline node in a network that supports at least one of cable television delivery and video delivery. One embodiment of the invention comprises a method having the steps of receiving the back channel to create a received back channel; partitioning the received back channel into a multiplicity of received identified speech channels; processing the multiplicity of received identified speech channels to create a multiplicity of identified speech content; and responding to the identified speech content to create an identified speech content response that is unique, for each of the multiplicity of identified speech contents.

66 Claims, 35 Drawing Sheets



U.S. PATENT DOCUMENTS

RE32,012	E		10/1985	Pirz et al. 381/43
4,633,499	A		12/1986	Nishioka et al. 381/43
4,653,045	A		3/1987	Stanley et al. 370/62
4,757,541	A		7/1988	Beadles 381/43
4,780,906	A		10/1988	Rajasekaran et al. 381/43
4,783,809	A		11/1988	Glinski 381/43
4,797,910	A		1/1989	Daudelin 379/67
4,821,027	A		4/1989	Mallory et al. 340/521
4,827,520	A		5/1989	Zeinstra 381/43
4,837,830	A	*	6/1989	Wrench et al. 704/238
4,847,885	A		7/1989	Vittorelli 379/6
4,907,079	A		3/1990	Turner et al. 358/84
4,931,950	A		6/1990	Isle et al. 364/513
4,959,855	A		9/1990	Daudelin 379/213
4,972,485	A		11/1990	Dautrich et al. 381/43
4,991,217	A		2/1991	Garrett et al. 381/43
5,006,987	A		4/1991	Harless 364/419
5,125,024	A		6/1992	Gokcen et al. 379/88
5,127,003	A		6/1992	Doll, Jr. et al. 370/110.1
5,155,760	A		10/1992	Johnson et al. 379/67
5,163,083	A		11/1992	Dowden et al. 379/88
5,165,095	A		11/1992	Borcherding 379/88
5,181,237	A		1/1993	Dowden et al. 379/88
5,197,005	A		3/1993	Shwartz et al. 364/419
5,199,062	A		3/1993	Von Meister et al. 379/67
5,208,848	A		5/1993	Pula 379/67
5,220,657	A	*	6/1993	Bly et al. 711/152
5,231,670	A		7/1993	Goldhor et al. 381/43
5,247,347	A		9/1993	Litteral et al. 358/85
5,255,305	A		10/1993	Sattar 379/34
5,271,088	A		12/1993	Bahler 395/2
5,274,560	A		12/1993	LaRue 364/444
5,325,421	A		6/1994	Hou et al. 379/67
5,329,608	A		7/1994	Bocchieri et al. 395/2.52
5,333,237	A		7/1994	Stefanopoulos et al. 395/12
5,335,276	A		8/1994	Thompson et al. 380/21
5,345,538	A		9/1994	Narayannan et al. 395/2.84
5,349,678	A		9/1994	Morris et al. 395/800
5,353,336	A		10/1994	Hou et al. 379/67
5,369,575	A		11/1994	Lamberti et al. 364/419.08
5,386,556	A		1/1995	Hedin et al. 395/600
5,434,777	A		7/1995	Luciw 364/419.13
5,442,749	A		8/1995	Northcutt et al. 395/200.09
5,446,489	A		8/1995	Egendorf 348/3
5,475,792	A		12/1995	Stanford et al. 395/2.42
5,500,920	A		3/1996	Kupiec 395/2.79
5,513,298	A		4/1996	Stanford et al. 395/2.52
5,519,608	A		5/1996	Kupiec 364/419.08
5,526,034	A		6/1996	Hoarty et al. 348/7
5,550,578	A		8/1996	Hoarty et al. 348/7
5,553,119	A		9/1996	McAllister et al. 379/67
5,564,001	A		10/1996	Lewis 395/154
5,566,271	A		10/1996	Tomitsuka et al. 395/2.84
5,585,858	A		12/1996	Harper et al. 348/485
5,602,963	A		2/1997	Bissonnette et al. 395/2.84
5,608,448	A		3/1997	Smoral et al. 348/7
5,608,624	A		3/1997	Luciw 395/794
5,615,296	A		3/1997	Stanford et al. 395/2.1
5,625,814	A		4/1997	Luciw 395/605
5,634,086	A		5/1997	Rtischev et al. 395/2.79
5,642,519	A		6/1997	Martin 395/759
5,664,061	A		9/1997	Andreshak et al. 704/275
5,666,400	A		9/1997	McAllister et al. 379/67
5,668,929	A		9/1997	Foster, Jr. 704/273
5,671,225	A		9/1997	Hooper et al. 370/468
5,671,328	A		9/1997	Fitzpatrick et al. 395/2.55
5,719,921	A		2/1998	Vysotsky et al. 379/88
5,721,827	A		2/1998	Logan et al. 395/200.47
5,721,902	A		2/1998	Schultz 395/604
5,721,938	A		2/1998	Stuckey 395/754
5,726,984	A		3/1998	Kubler et al. 370/349
5,729,659	A		3/1998	Potter 395/2.79
5,732,216	A		3/1998	Logan et al. 395/200.33
5,737,495	A		4/1998	Adams et al. 395/615
5,748,841	A		5/1998	Morin et al. 395/2.66
5,748,974	A		5/1998	Johnson 395/759
5,752,232	A		5/1998	Basore et al. 704/275
5,752,246	A		5/1998	Rogers et al. 707/10
5,774,525	A		6/1998	Kanevsky et al. 379/88
5,774,841	A		6/1998	Salazar et al. 704/225
5,774,859	A		6/1998	Houser et al. 704/275
5,774,860	A		6/1998	Bayya et al. 704/275
5,777,571	A		7/1998	Chuang 341/176
5,794,050	A		8/1998	Dahlgren et al. 395/708
5,794,205	A		8/1998	Walters et al. 704/275
5,802,526	A		9/1998	Fawcett et al. 707/104
5,805,775	A		9/1998	Eberman et al. 395/12
5,809,471	A		9/1998	Brodsky 704/275
5,812,977	A		9/1998	Douglas 704/275
5,819,220	A		10/1998	Sarukkai et al. 704/243
5,819,225	A		10/1998	Eastwood et al. 704/275
5,821,508	A		10/1998	Willard 235/51
5,822,727	A		10/1998	Garberg et al. 701/243
5,826,267	A		10/1998	McMillan 707/9
5,832,063	A		11/1998	Vysotsky et al. 379/88
5,832,439	A		11/1998	Cox, Jr. et al. 704/275
5,855,002	A		12/1998	Armstrong 704/270
5,860,059	A		1/1999	Aust et al. 704/231
5,864,815	A		1/1999	Rozak et al. 704/275
5,864,819	A		1/1999	De Armas et al. 704/275
5,867,817	A		2/1999	Catallo et al. 704/255
5,874,939	A		2/1999	Galvin 345/156
5,884,265	A		3/1999	Squitteri et al. 704/275
5,885,083	A		3/1999	Ferrell 434/156
5,890,122	A		3/1999	Van Kleeck et al. 704/275
5,890,123	A		3/1999	Brown et al. 704/275
5,892,813	A		4/1999	Morin et al. 379/88.01
5,893,063	A		4/1999	Loats et al. 704/275
5,903,870	A		5/1999	Kaufman 704/275
5,915,001	A		6/1999	Uppaluru 379/88.22
5,917,891	A		6/1999	Will 379/88.03
5,920,841	A		7/1999	Schottmuller et al. 704/275
5,923,738	A		7/1999	Cardillo, IV et al. 379/93.25
5,929,850	A		7/1999	Broadwin et al. 345/327
5,930,341	A		7/1999	Cardillo, IV et al. 379/93.25
5,933,835	A		8/1999	Adams et al. 707/104
5,937,041	A		8/1999	Cardillo, IV et al. 379/93.25
5,946,050	A		8/1999	Wolff 348/553
5,953,700	A		9/1999	Kanevsky et al. 704/246
5,960,399	A		9/1999	Barclay et al. 704/270
5,963,940	A		10/1999	Liddy et al. 707/5
5,970,457	A		10/1999	Brant et al. 704/275
5,974,385	A		10/1999	Ponting et al. 704/275
5,991,365	A		11/1999	Pizano et al. 379/88.13
5,991,719	A		11/1999	Yazaki et al. 704/251
5,999,970	A		12/1999	Krisbergh et al. 709/217
6,003,072	A		12/1999	Gerritsen et al. 709/218
6,009,398	A		12/1999	Mueller et al. 704/275
6,009,470	A		12/1999	Watkins 709/231
6,012,030	A		1/2000	French-St. George et al. 704/275
6,018,711	A		1/2000	French-St. George et al. 704/275
6,021,371	A		2/2000	Fultz 701/200
6,026,388	A		2/2000	Liddy et al. 707/1
6,034,678	A		3/2000	Hoarty et al. 345/327
6,035,267	A		3/2000	Watanabe et al. 704/1
6,035,275	A		3/2000	Brode et al. 704/275
6,084,572	A		7/2000	Yaniger et al. 345/159
6,100,883	A		8/2000	Hoarty 345/327
6,173,279	B1		1/2001	Levin et al. 707/5
6,181,334	B1		1/2001	Freeman et al. 345/327
6,192,338	B1		2/2001	Haszto et al. 704/257
6,198,822	B1		3/2001	Doyle et al. 380/210

6,204,843	B1	3/2001	Freemen et al.	345/327
6,205,582	B1	3/2001	Hoarty	725/93
6,215,484	B1	4/2001	Freemen et al.	345/327
6,225,976	B1	5/2001	Yates et al.	345/156
6,480,703	B1 *	11/2002	Calderone et al.	455/118
6,573,907	B1 *	6/2003	Madrane	715/719
6,785,653	B1 *	8/2004	White et al.	704/270.1

FOREIGN PATENT DOCUMENTS

WO	98/26595	6/1998	7/173
WO	99/30501	6/1999	7/173
WO	00/24198	4/2000	7/173

OTHER PUBLICATIONS

Sprint Begins Voice-Activated E-Mail; May 23, 1001; Associated Press-New York Times.

John Dowding, Jean Mark Gawron, Doug Appelt, John Bear, Lynn Cherny, Robert Moore, and Douglas Moran; *Gemini: a Natural Language System for Spoken-Language Understanding*.

Info Wiz: An Animated Voice Interactive Information System; http://www.ai.sri.com/~oaa/infowiz.html.

Robert Moore, Douglas Appelt, John Dowding, J. Mark Gawron, and Douglas Moran; *Combining Linguistic and Statistical Knowledge Sources in Natural-Language Processing for ATIS*.

Amanda Stent, John Dowding, Jean Mark Gawron, eleizabeth Owen Bratt, and Robert Moore; *The CommandTalk Spoken Dialogue System*.

John Dowding, Elizabeth Owen Bratt and Sharon Goldwater;*Interpreting Language in Context in CommandTalk* Feb. 5, 1999.

Robert Moore, John Dowding, Harry Bratt, J. Mark Gawron, Yonael Gorfu, and Adam Cheyer; *CommandTalk: A Spoken-Language Interface for Battlefield Simulations*; Oct. 23, 1997.

John Dowding, Robert Moore, Francois Andry, and Douglas Moran; *Interleaving Syntax and Semantics in an Efficient Bottom-Up Parser*.

Robert C. Bedichek; *The Meerkat Multicomputer: Tradeoffs in Multicomputer Architecture*; Dissertatioin Given and University of Washington, 1994.

James Karney; *A Sound Proposition Even Early on, There are Already Ways to Voice-enable a Site*; Jan. 15, 2001; Internet World.

Sarah Roberts-Witt; *Talking to the Web*; Jan. 15, 2001; Internet World.

Hand out from SRI International; *Simple Ui to Retrieve Favorites*.

Email dated Jan. 15, 2001; Subject *Fwd: Large Linux Cluster*.

Email: *Cray inc. Announces Plans for Alpha Linux Supercluster System*; Feb. 12, 2001.

Email: *Fwd: Re: One Voice Technologies*; Feb. 15, 2001.

Email: *Fwd: Ars Technica: Inside the KLAT2 Supercomputer--p. 1-*.

53 Page printout from: http://www.ai.sri.com CommandTalk.

Printout: http://www.speech.sri.com/demos/atis.html.

Bruce Meyerson. *Vice Recognition Still Selective*, Jan. 15, 2000, Yahoo! News Alerts.

Where can I get an inVoca?, http://invoca.com/home/htm.

Integrated Audio-Graphics User Interface., Apr. 1991, IBM Technical Disclosure Bulletin.

Stephanie Miles, *Will Microsoft's next OS run your TV?*, Sep. 5, 2000, Yahoo! Daily News.

* cited by examiner

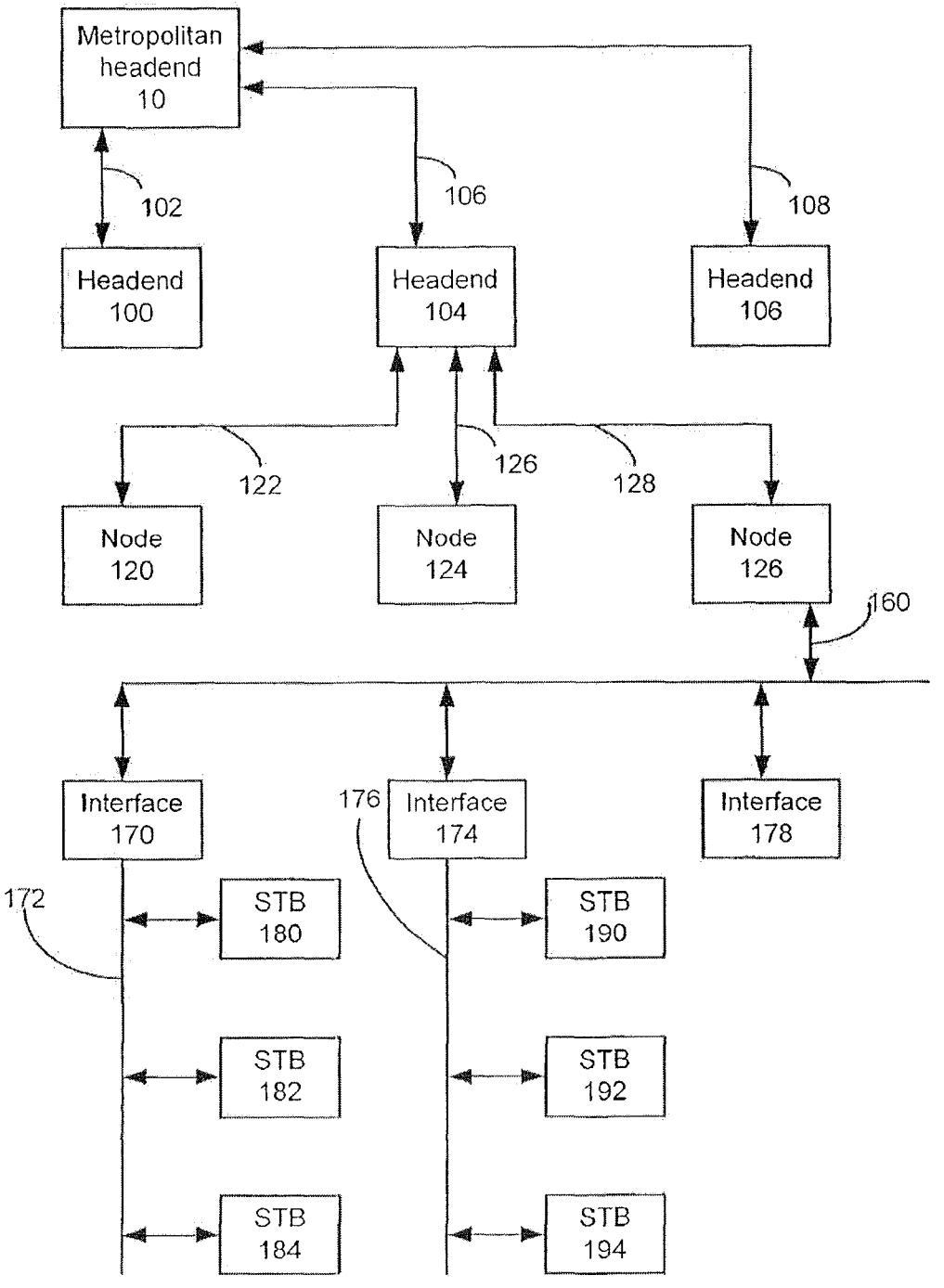


Fig. 1 Prior Art

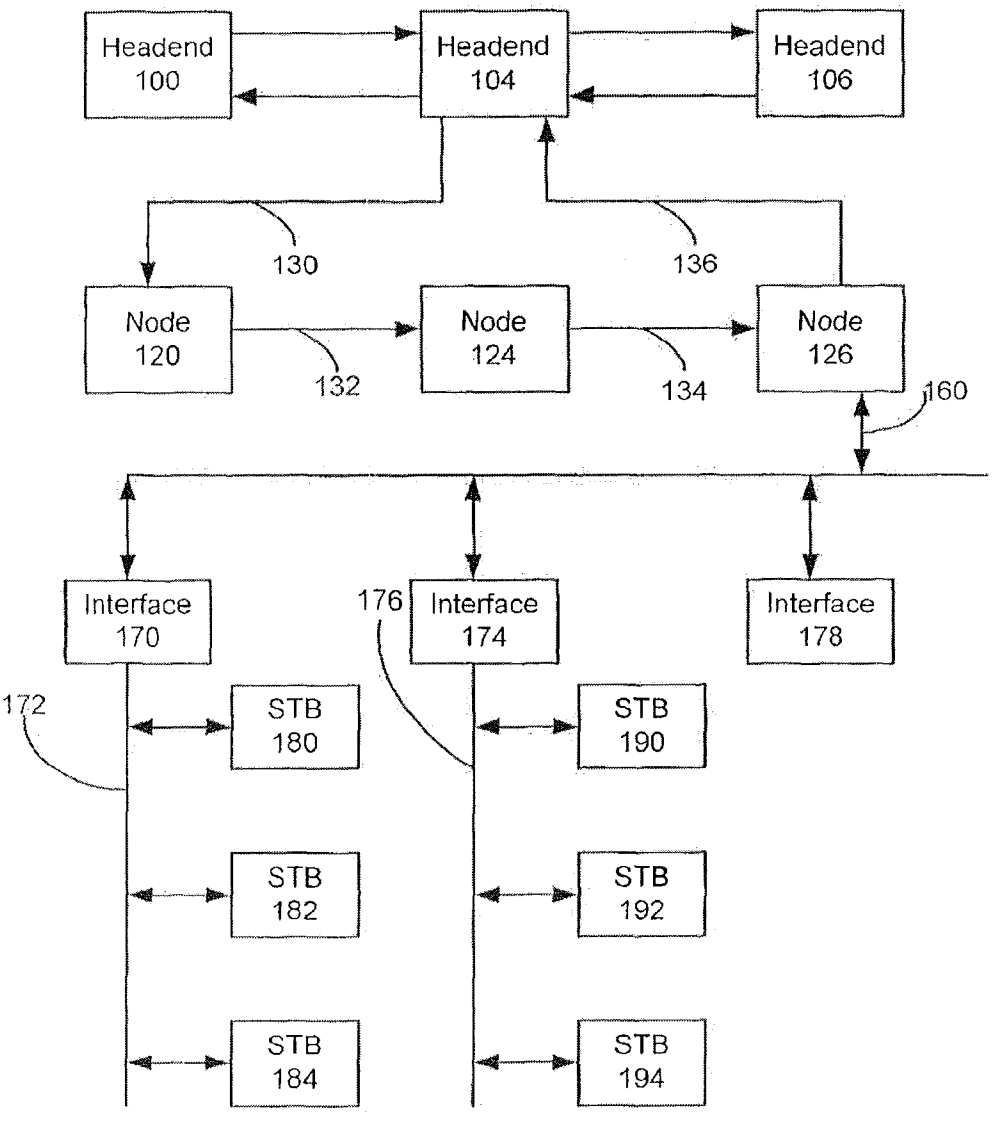


Fig. 2 Prior Art

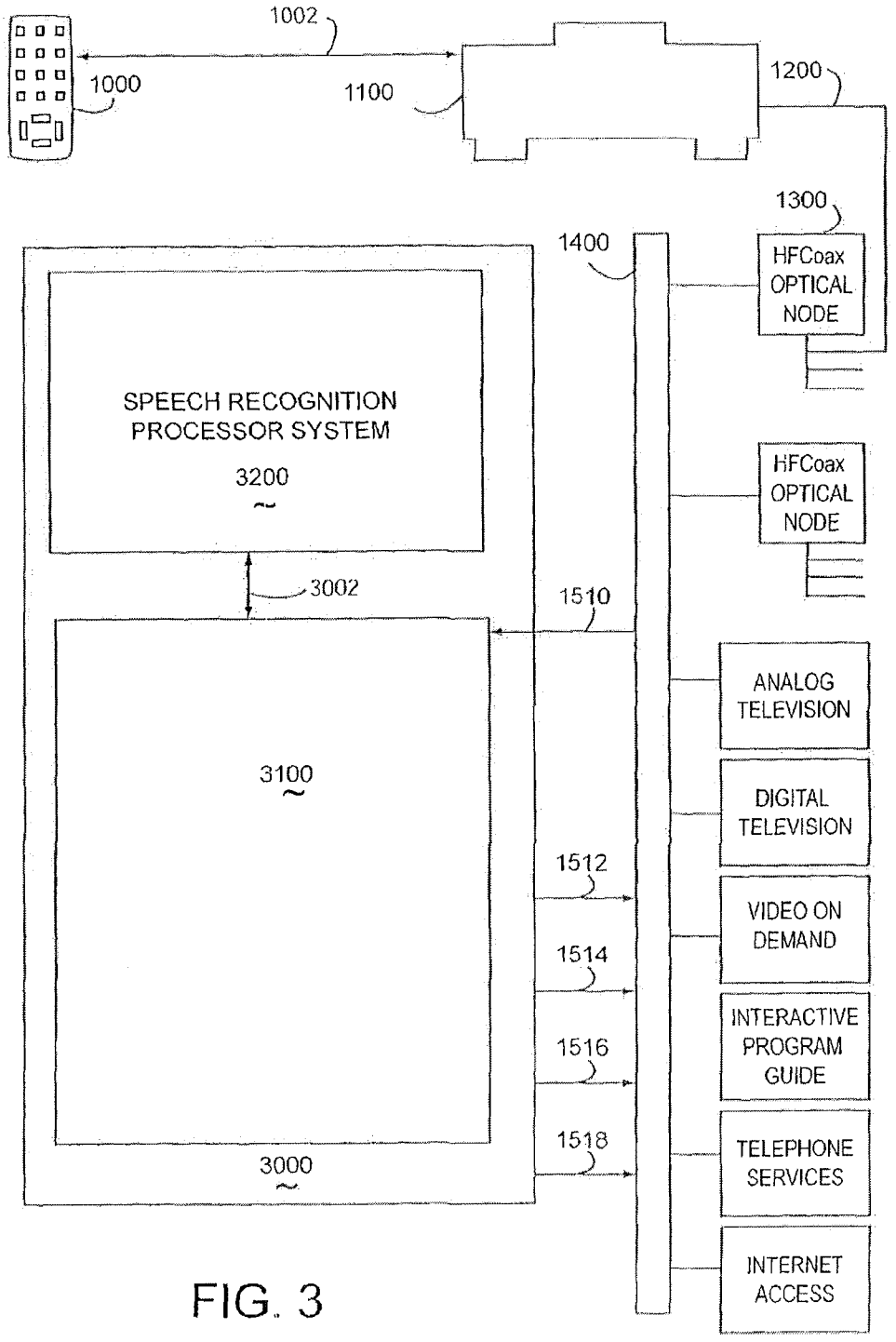

1002
1000
1100
1200
1300
1400
HFCoax OPTICAL NODE
HFCoax OPTICAL NODE
SPEECH RECOGNITION PROCESSOR SYSTEM
3200
3002
1510
3100
ANALOG TELEVISION
DIGITAL TELEVISION
1512
VIDEO ON DEMAND
1514
INTERACTIVE PROGRAM GUIDE
1516
1518
TELEPHONE SERVICES
3000
INTERNET ACCESS

FIG. 3

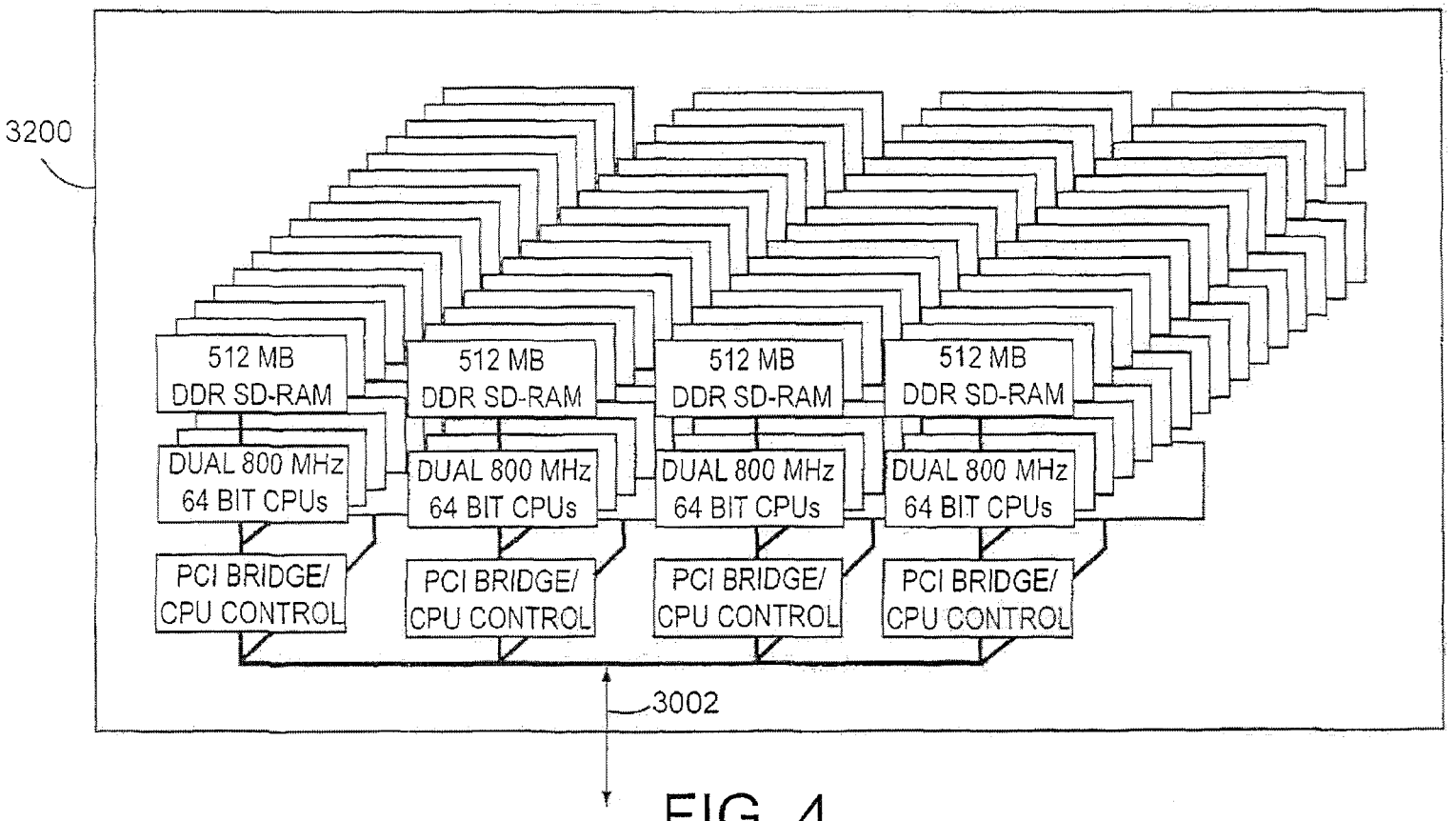

3200
512 MB DDR SD-RAM
512 MB DDR SD-RAM
512 MB DDR SD-RAM
512 MB DDR SD-RAM
DUAL 800 MHz 64 BIT CPUs
DUAL 800 MHz 64 BIT CPUs
DUAL 800 MHz 64 BIT CPUs
DUAL 800 MHz 64 BIT CPUs
PCI BRIDGE/ CPU CONTROL
PCI BRIDGE/ CPU CONTROL
PCI BRIDGE/ CPU CONTROL
PCI BRIDGE/ CPU CONTROL
3002

FIG. 4

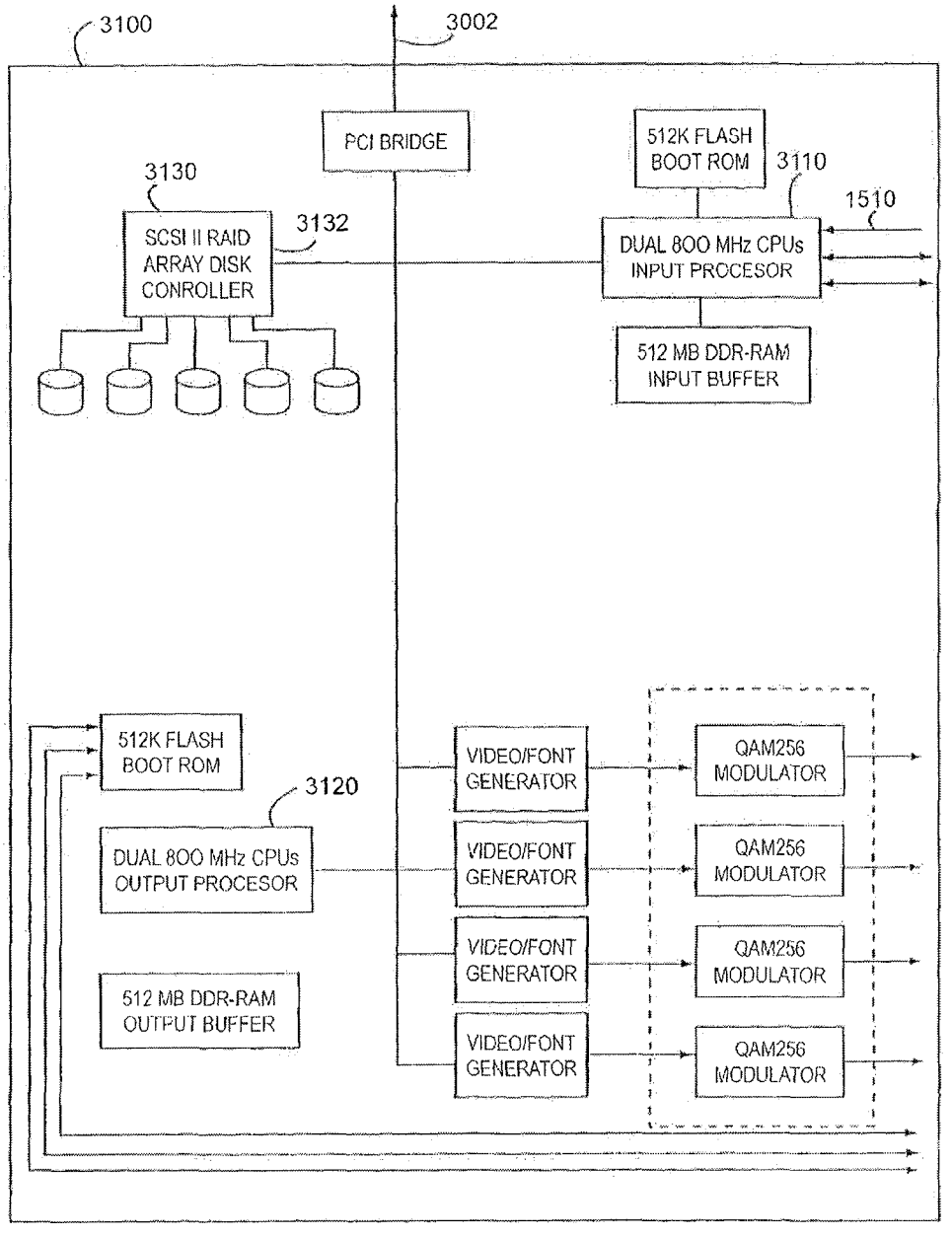

3100
3002
PCI BRIDGE
512K FLASH BOOT ROM
3110
1510
3130
SCSI II RAID ARRAY DISK CONROLLER
3132
DUAL 800 MHz CPUs INPUT PROCESOR
512 MB DDR-RAM INPUT BUFFER
512K FLASH BOOT ROM
3120
DUAL 800 MHz CPUs OUTPUT PROCESOR
512 MB DDR-RAM OUTPUT BUFFER
VIDEO/FONT GENERATOR
VIDEO/FONT GENERATOR
VIDEO/FONT GENERATOR
VIDEO/FONT GENERATOR
QAM256 MODULATOR
QAM256 MODULATOR
QAM256 MODULATOR
QAM256 MODULATOR

FIG. 5

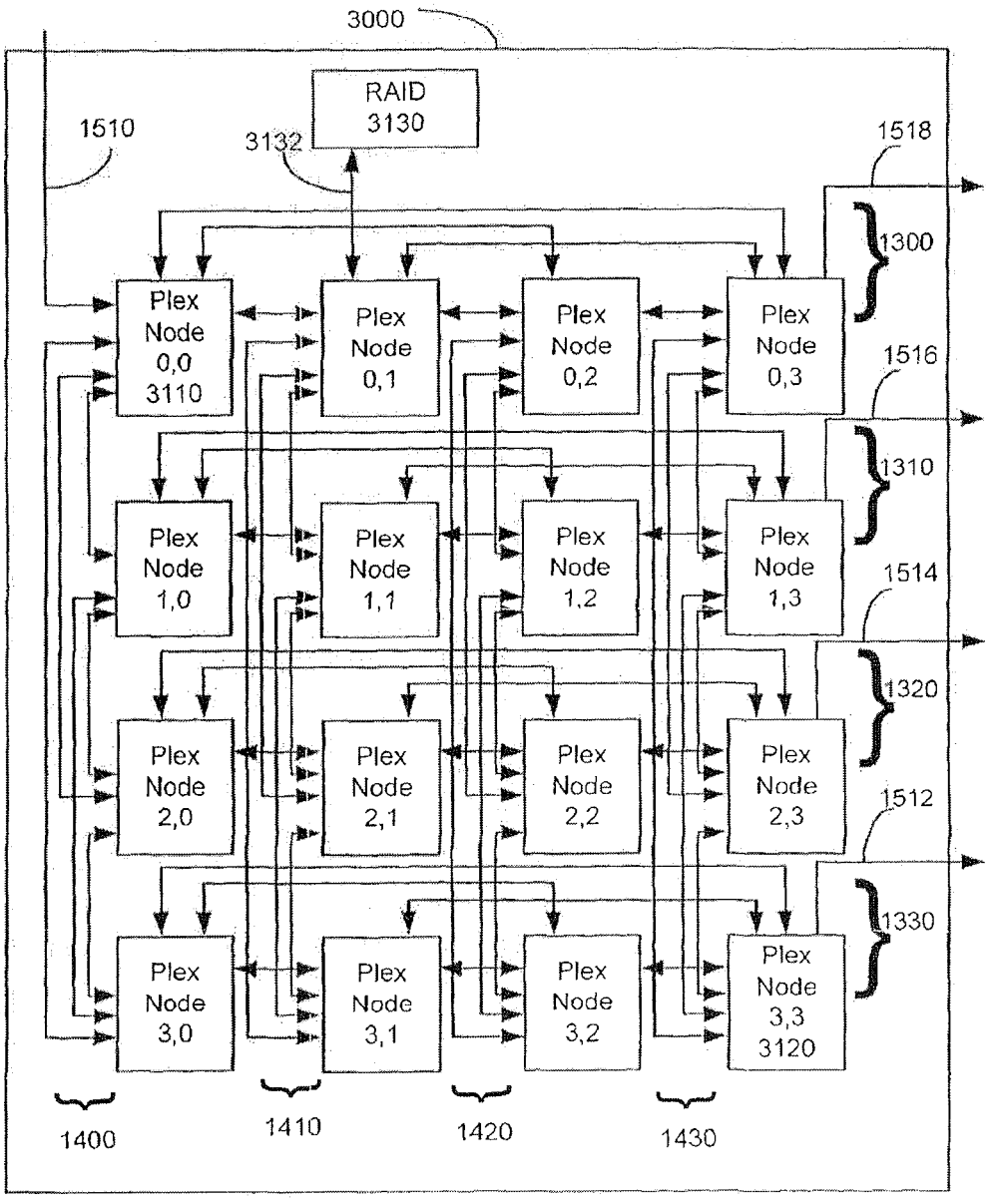

3000
RAID
3130
1510
3132
1518
1300
Plex
Node
0,0
3110
Plex
Node
0,1
Plex
Node
0,2
Plex
Node
0,3
1516
1310
Plex
Node
1,0
Plex
Node
1,1
Plex
Node
1,2
Plex
Node
1,3
1514
1320
Plex
Node
2,0
Plex
Node
2,1
Plex
Node
2,2
Plex
Node
2,3
1512
1330
Plex
Node
3,0
Plex
Node
3,1
Plex
Node
3,2
Plex
Node
3,3
3120
1400
1410
1420
1430

Fig. 6

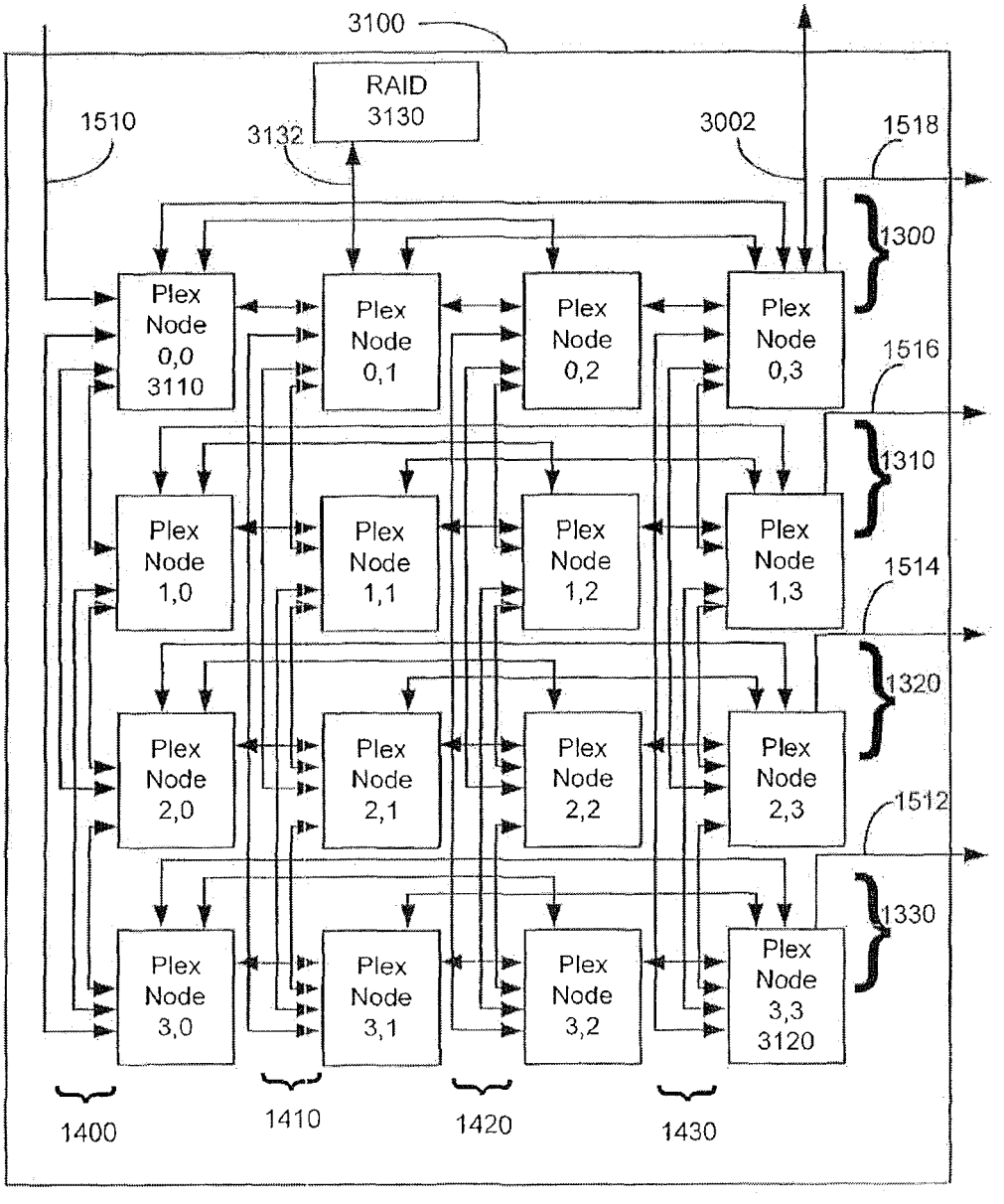

3100
RAID
3130
1510
3132
3002
1518
1300
Plex Node 0,0 3110
Plex Node 0,1
Plex Node 0,2
Plex Node 0,3
1516
1310
Plex Node 1,0
Plex Node 1,1
Plex Node 1,2
Plex Node 1,3
1514
1320
Plex Node 2,0
Plex Node 2,1
Plex Node 2,2
Plex Node 2,3
1512
1330
Plex Node 3,0
Plex Node 3,1
Plex Node 3,2
Plex Node 3,3 3120
1400
1410
1420
1430

Fig. 7

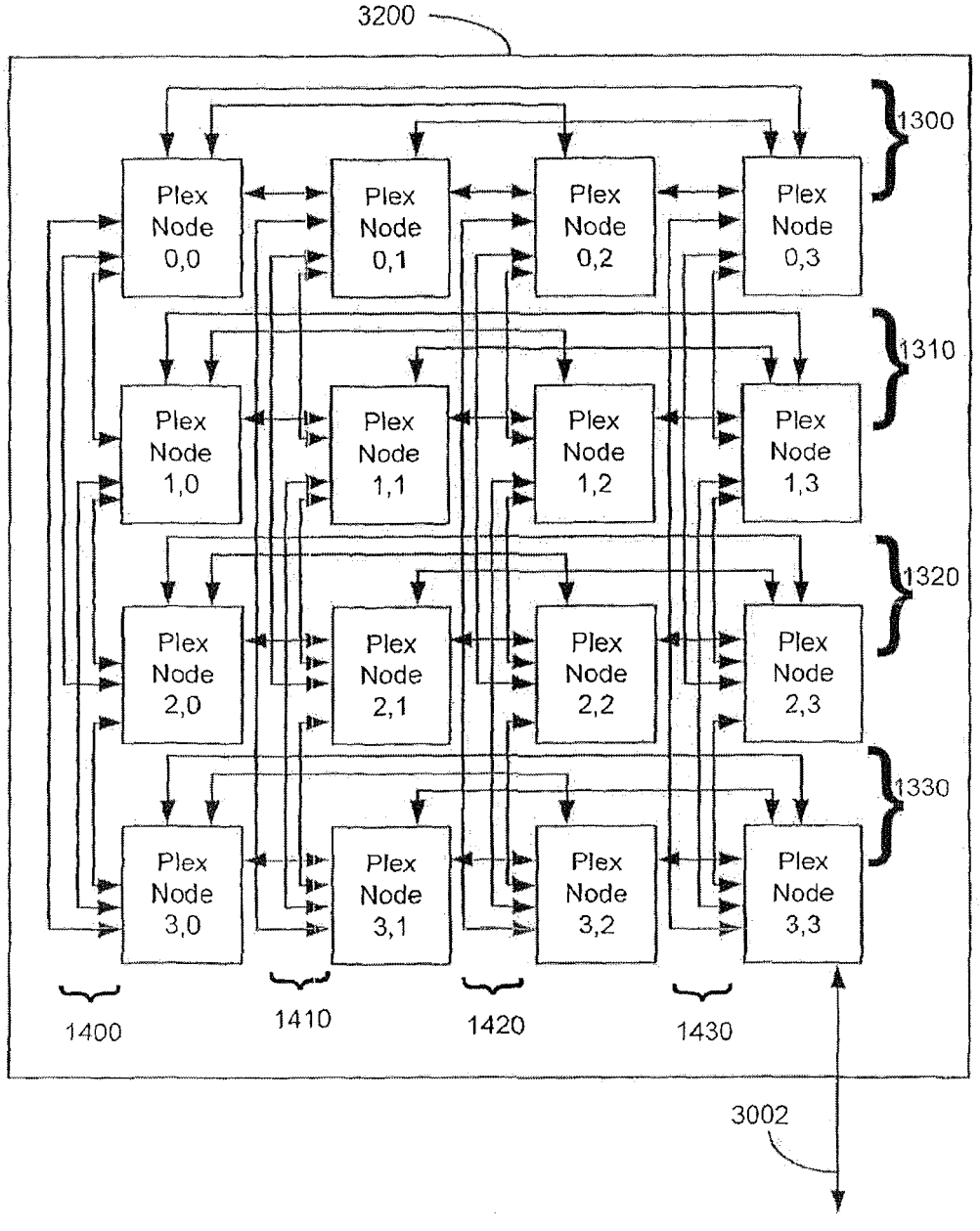

3200
1300
1310
1320
1330
Plex Node 0,0
Plex Node 0,1
Plex Node 0,2
Plex Node 0,3
Plex Node 1,0
Plex Node 1,1
Plex Node 1,2
Plex Node 1,3
Plex Node 2,0
Plex Node 2,1
Plex Node 2,2
Plex Node 2,3
Plex Node 3,0
Plex Node 3,1
Plex Node 3,2
Plex Node 3,3
1400
1410
1420
1430
3002

Fig. 8

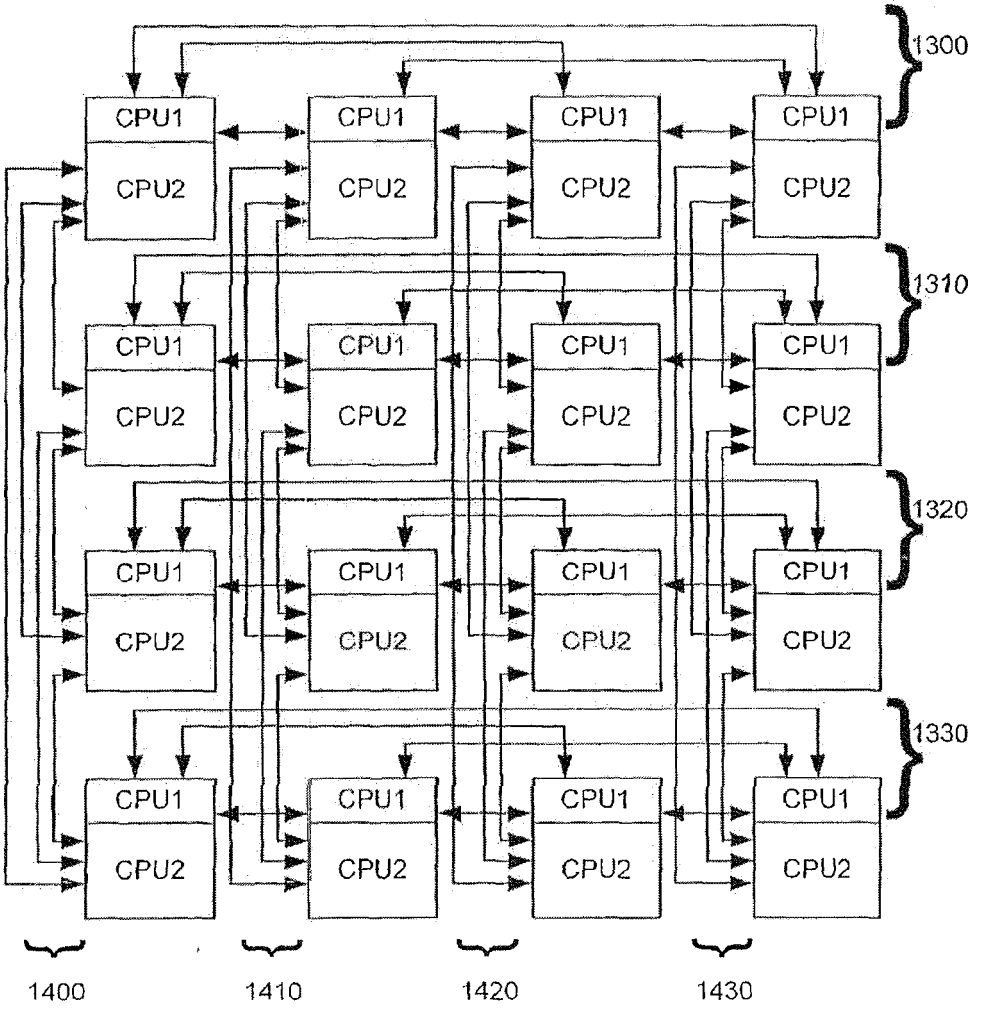


Fig. 9

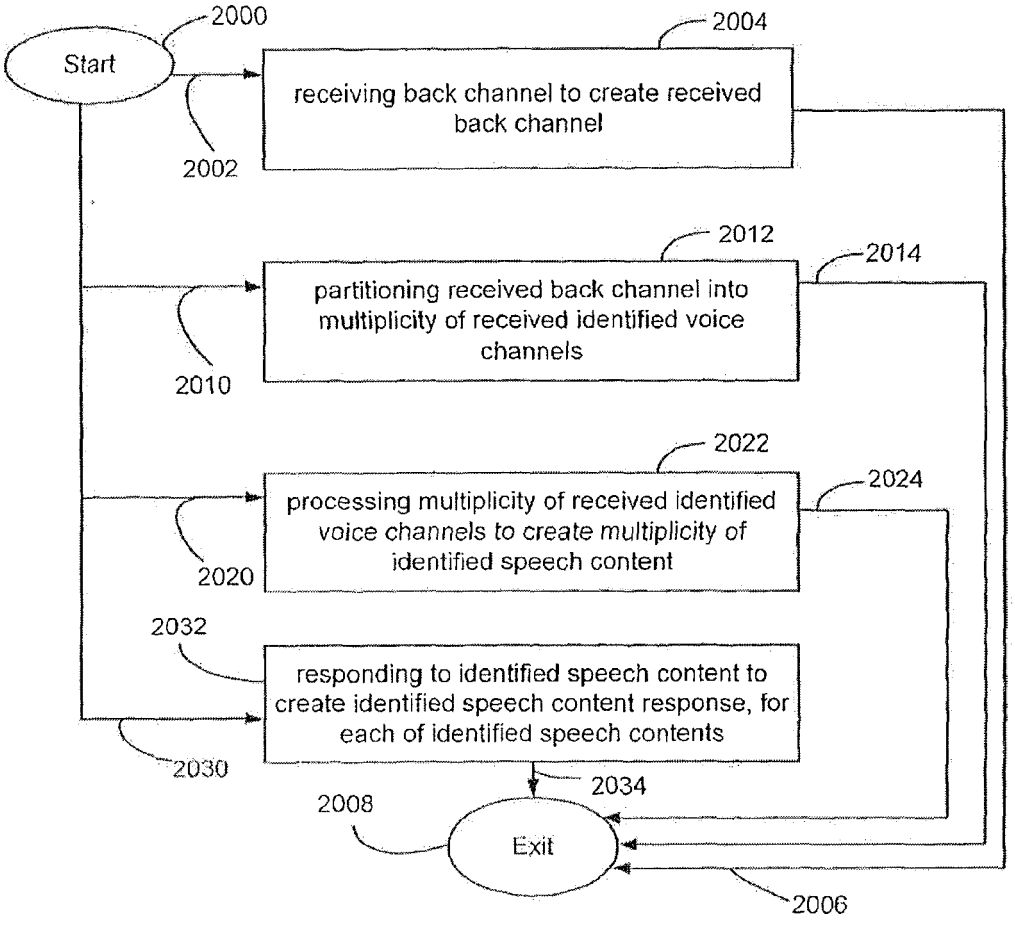

2000
Start
2004
receiving back channel to create received back channel
2002
2012
2014
partitioning received back channel into multiplicity of received identified voice channels
2010
2022
2024
processing multiplicity of received identified voice channels to create multiplicity of identified speech content
2020
2032
responding to identified speech content to create identified speech content response, for each of identified speech contents
2030
2034
2008
Exit
2006

Fig. 10

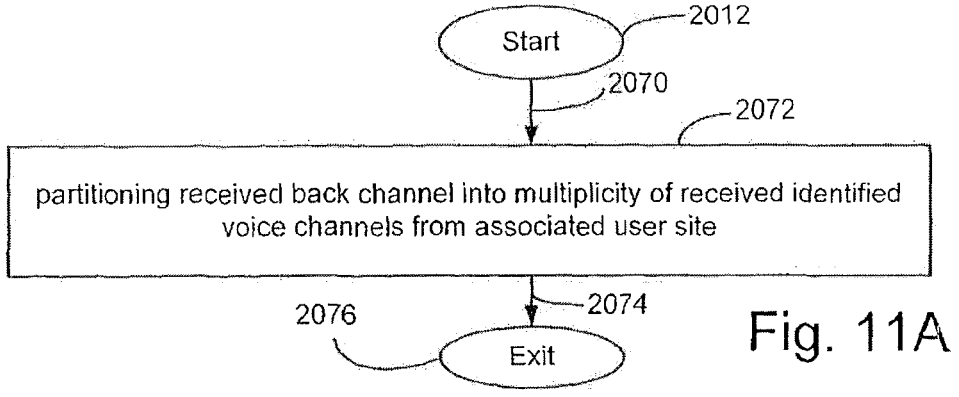


Fig. 11A

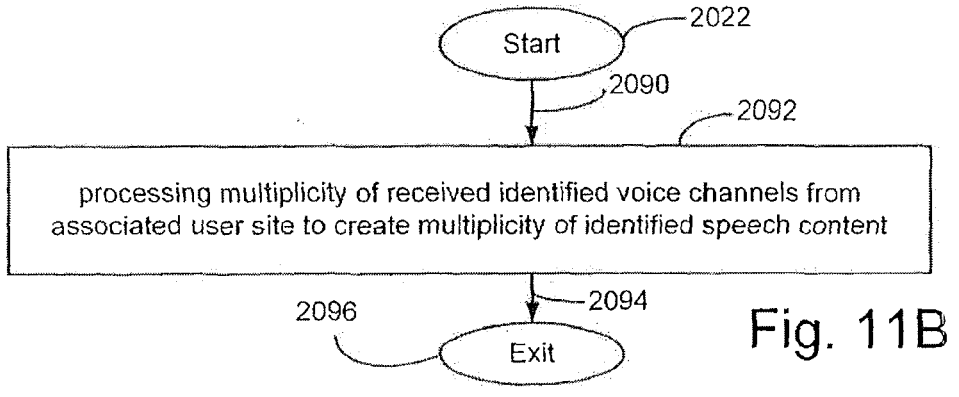


Fig. 11B

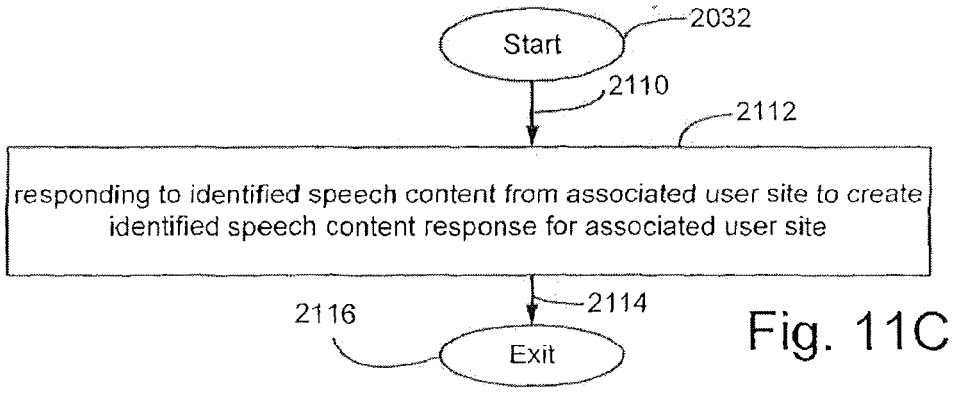


Fig. 11C

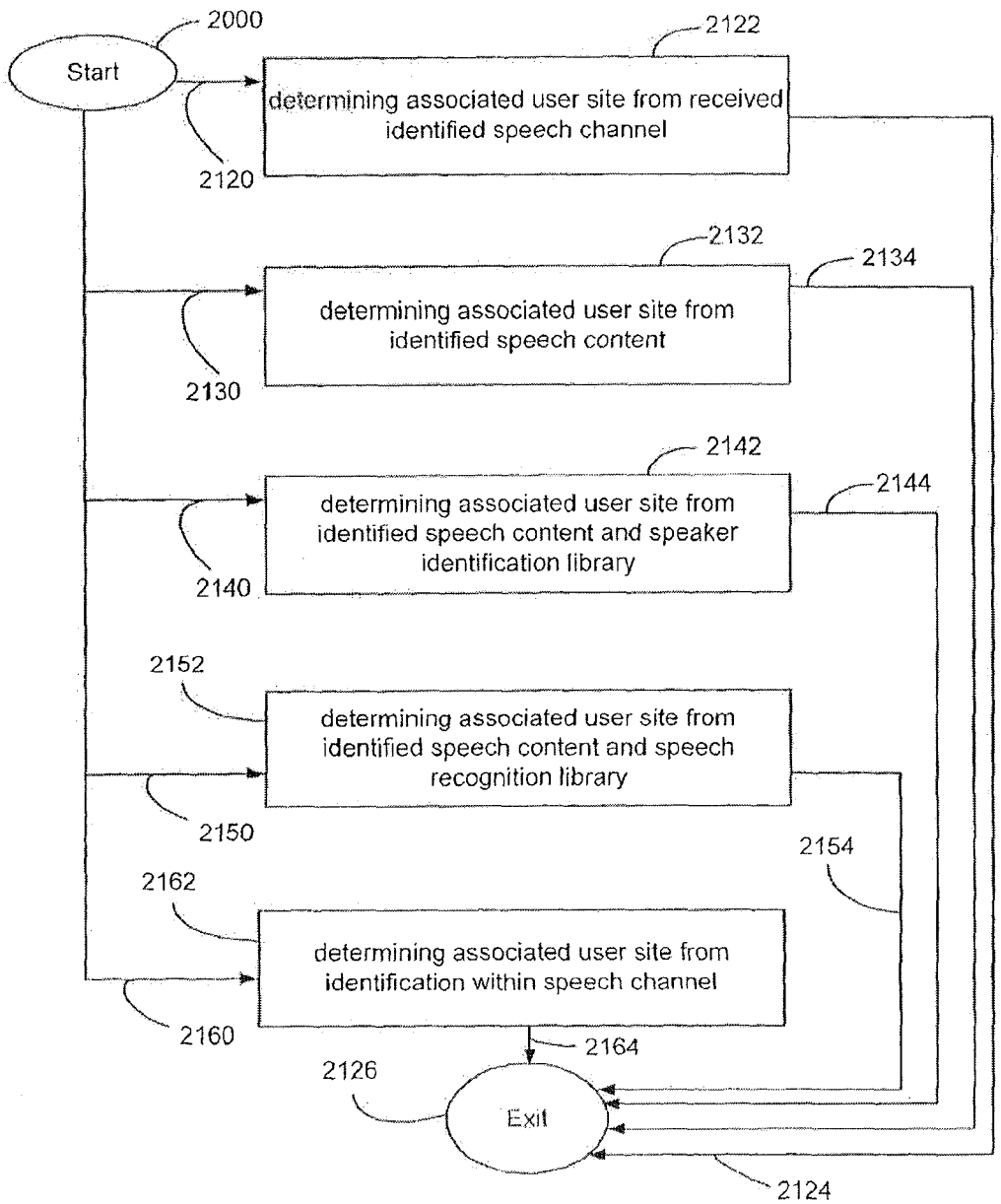

2000
Start
2122
determining associated user site from received identified speech channel
2120
2132
2134
determining associated user site from identified speech content
2130
2142
2144
determining associated user site from identified speech content and speaker identification library
2140
2152
determining associated user site from identified speech content and speech recognition library
2150
2154
2162
determining associated user site from identification within speech channel
2160
2164
2126
Exit
2124

Fig. 12

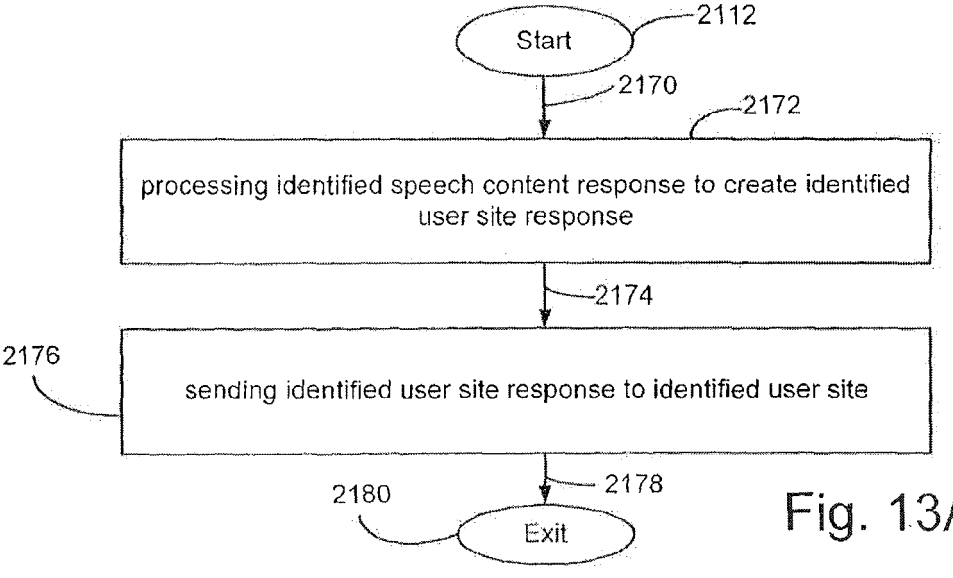


Fig. 13A

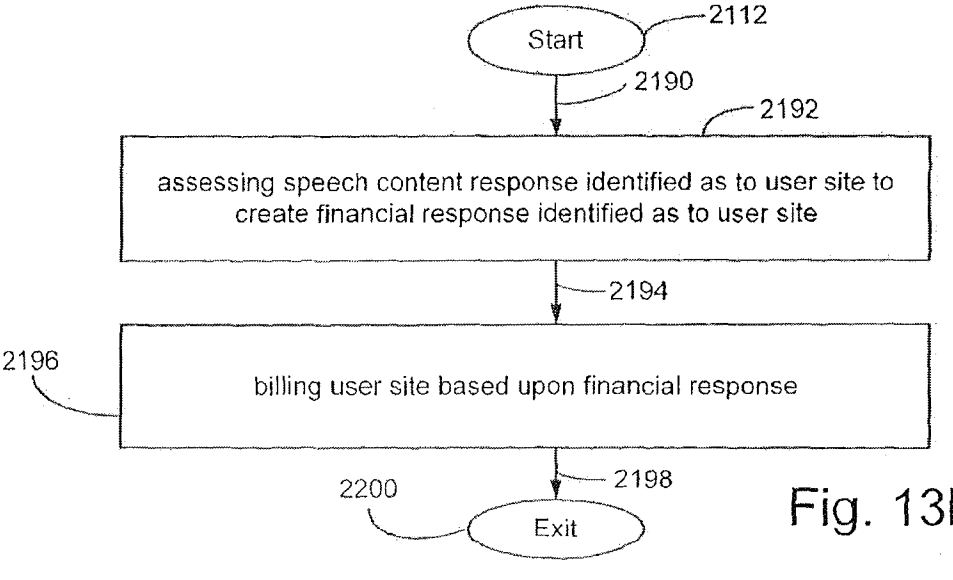


Fig. 13B

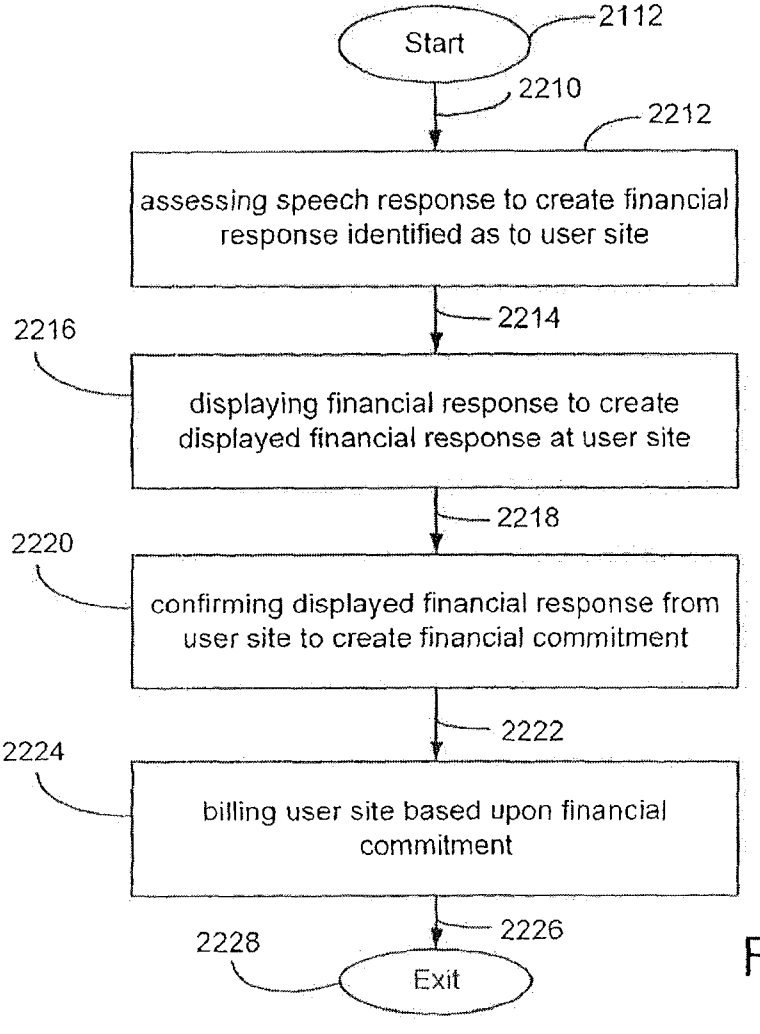

Start
2112
2210
2212
assessing speech response to create financial response identified as to user site
2214
2216
displaying financial response to create displayed financial response at user site
2218
2220
confirming displayed financial response from user site to create financial commitment
2222
2224
billing user site based upon financial commitment
2226
2228
Exit

Fig. 14

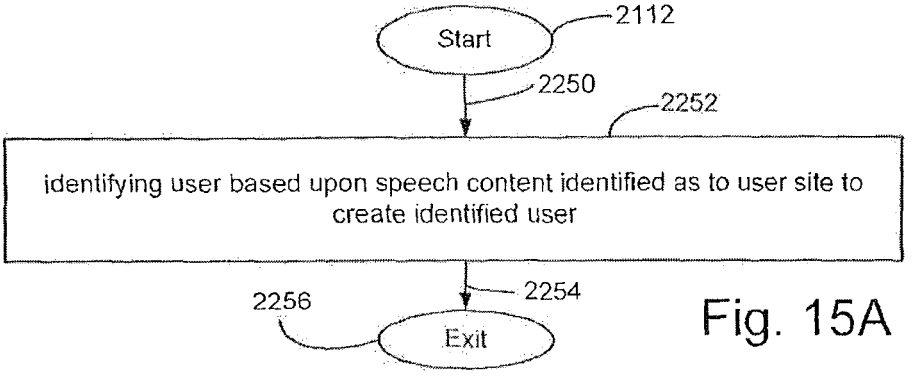


Fig. 15A

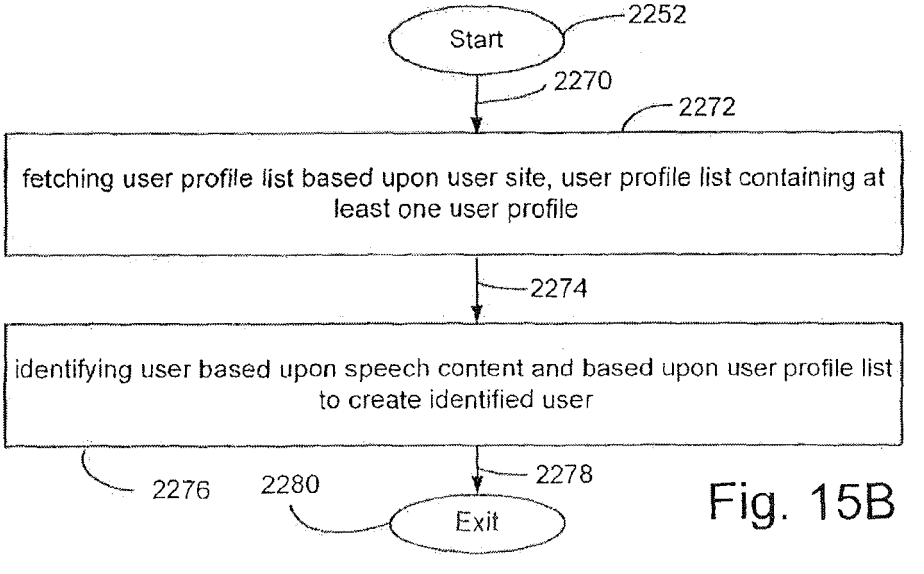


Fig. 15B

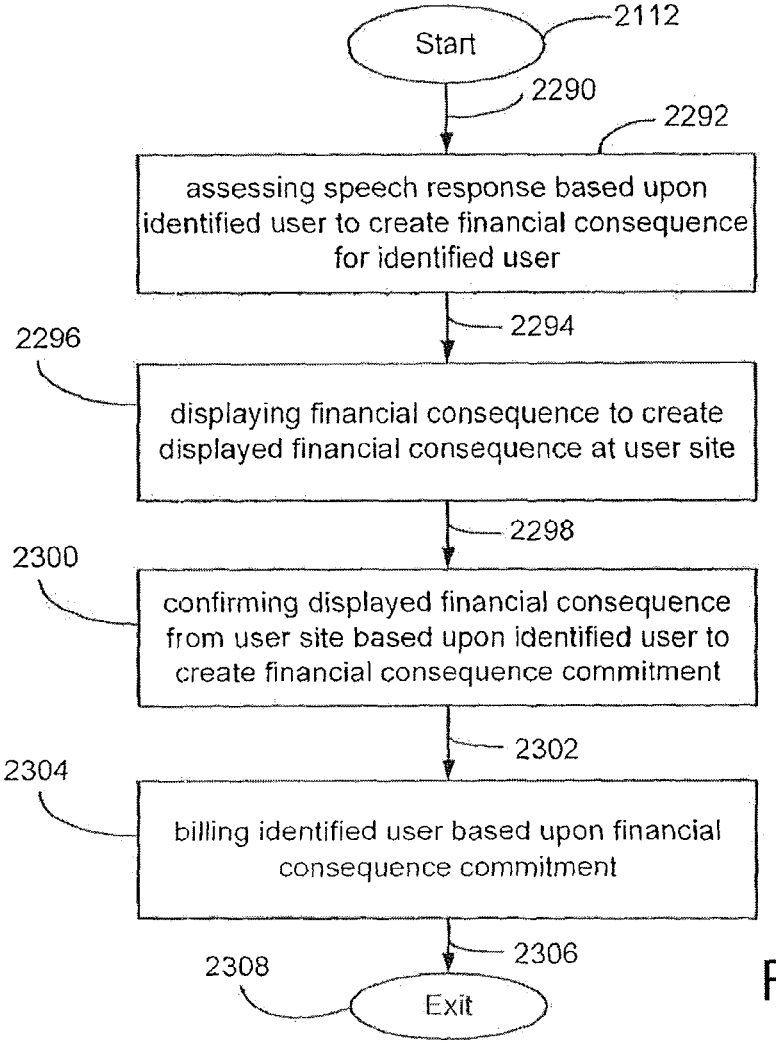

Start
2112
2290
2292
assessing speech response based upon identified user to create financial consequence for identified user
2294
2296
displaying financial consequence to create displayed financial consequence at user site
2298
2300
confirming displayed financial consequence from user site based upon identified user to create financial consequence commitment
2302
2304
billing identified user based upon financial consequence commitment
2306
2308
Exit

Fig. 16

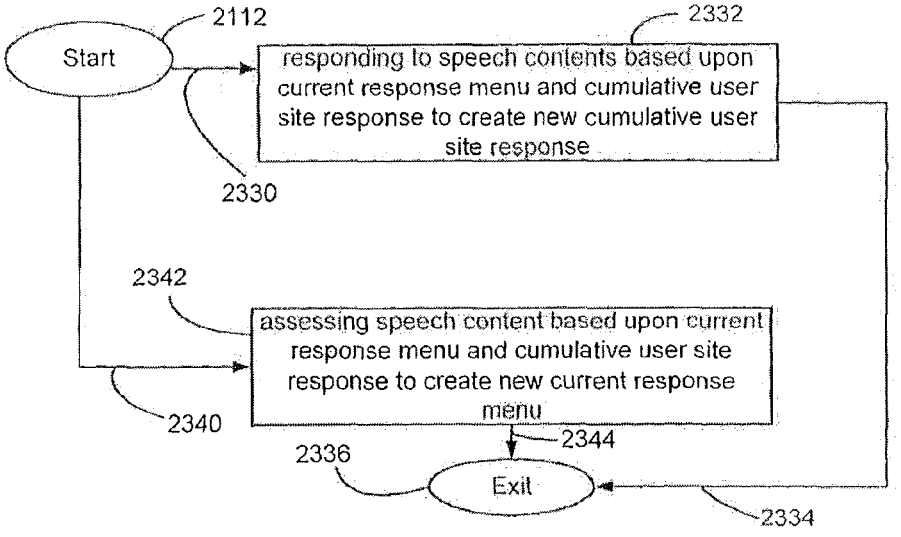


Fig. 17B

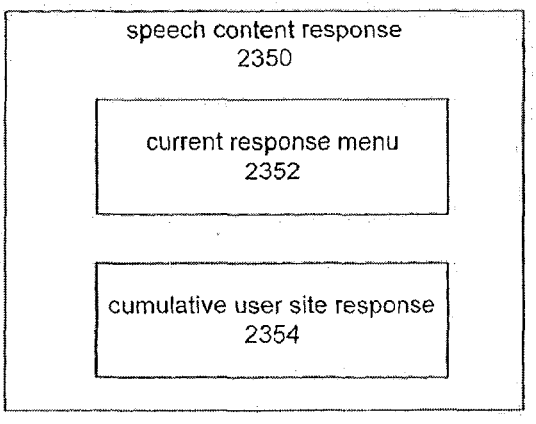


Fig. 17A

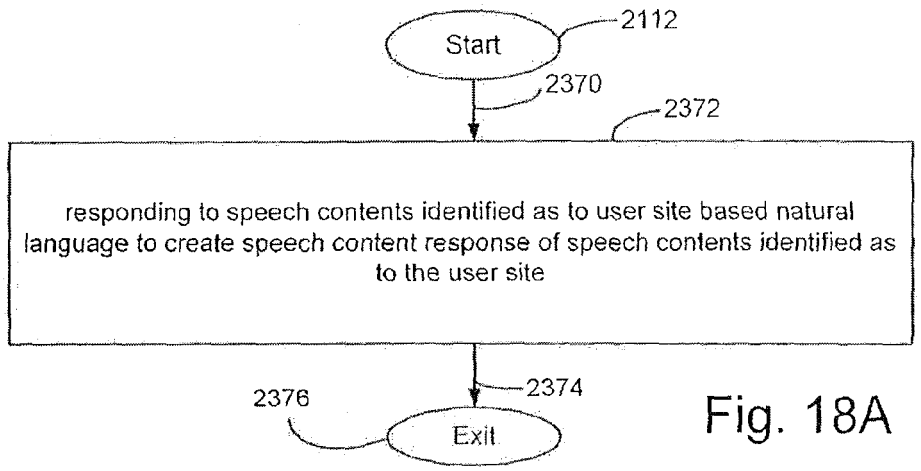


Fig. 18A

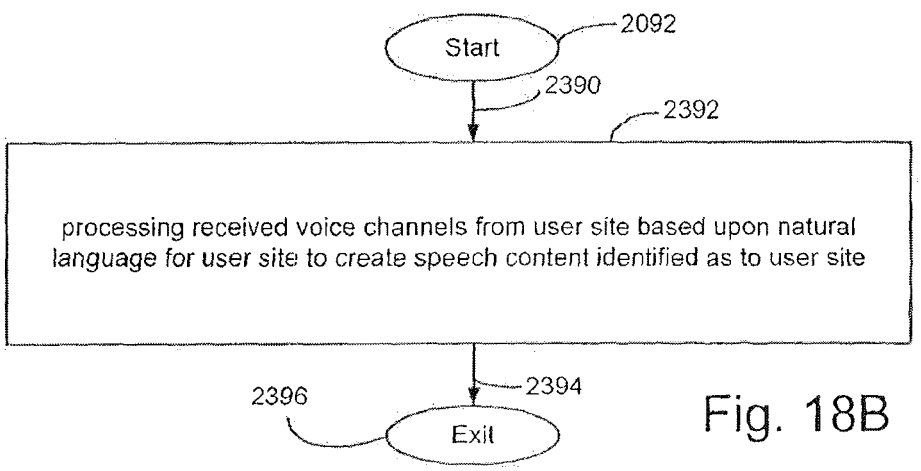


Fig. 18B

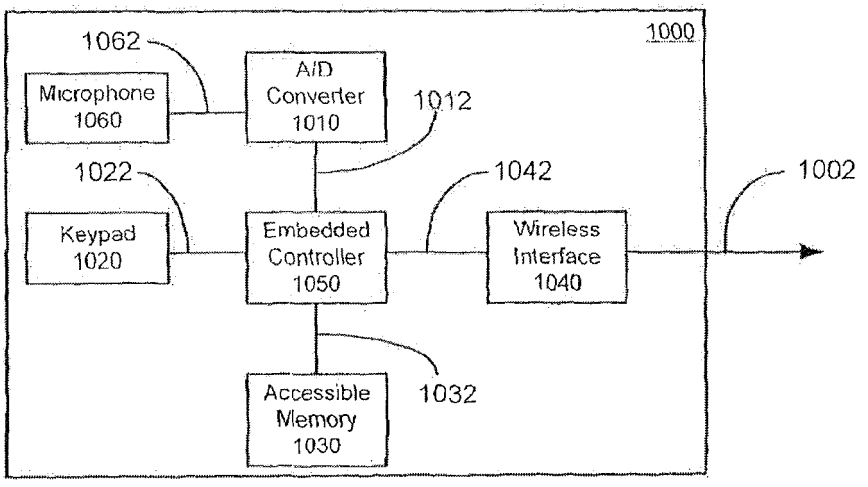


Fig. 19A

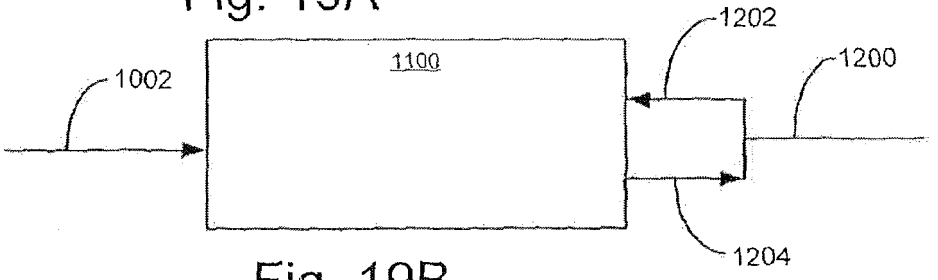


Fig. 19B

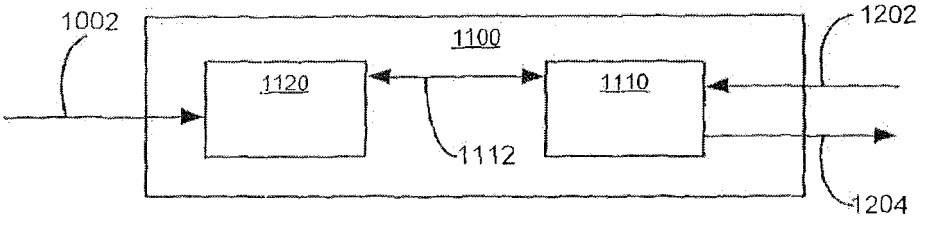


Fig. 19C

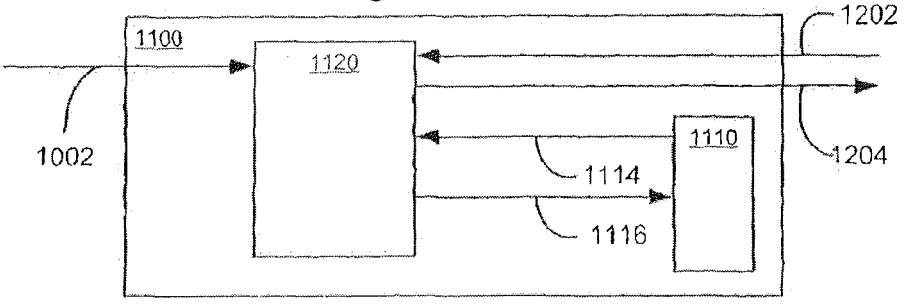


Fig. 19D

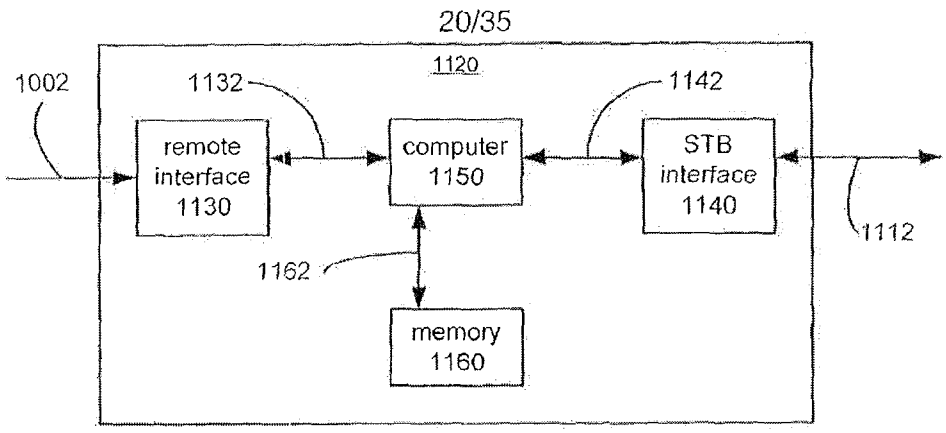


Fig. 20A

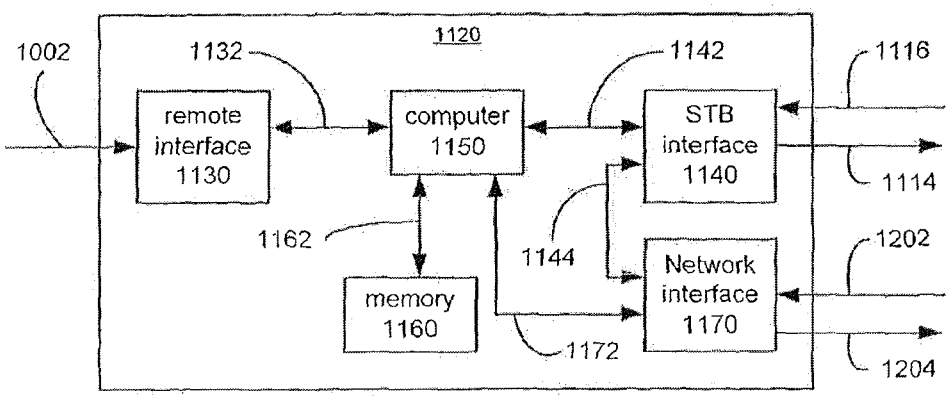


Fig. 20B

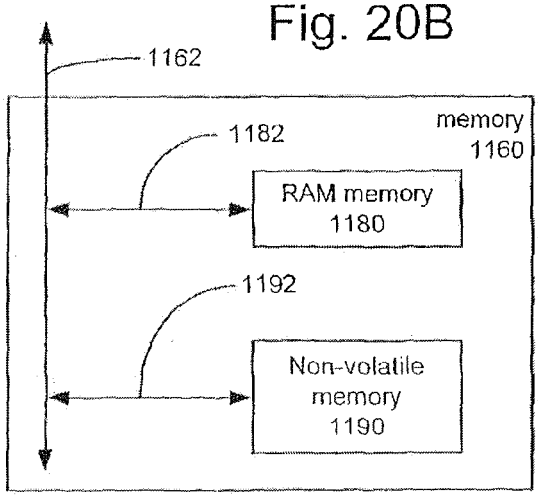


Fig. 20C

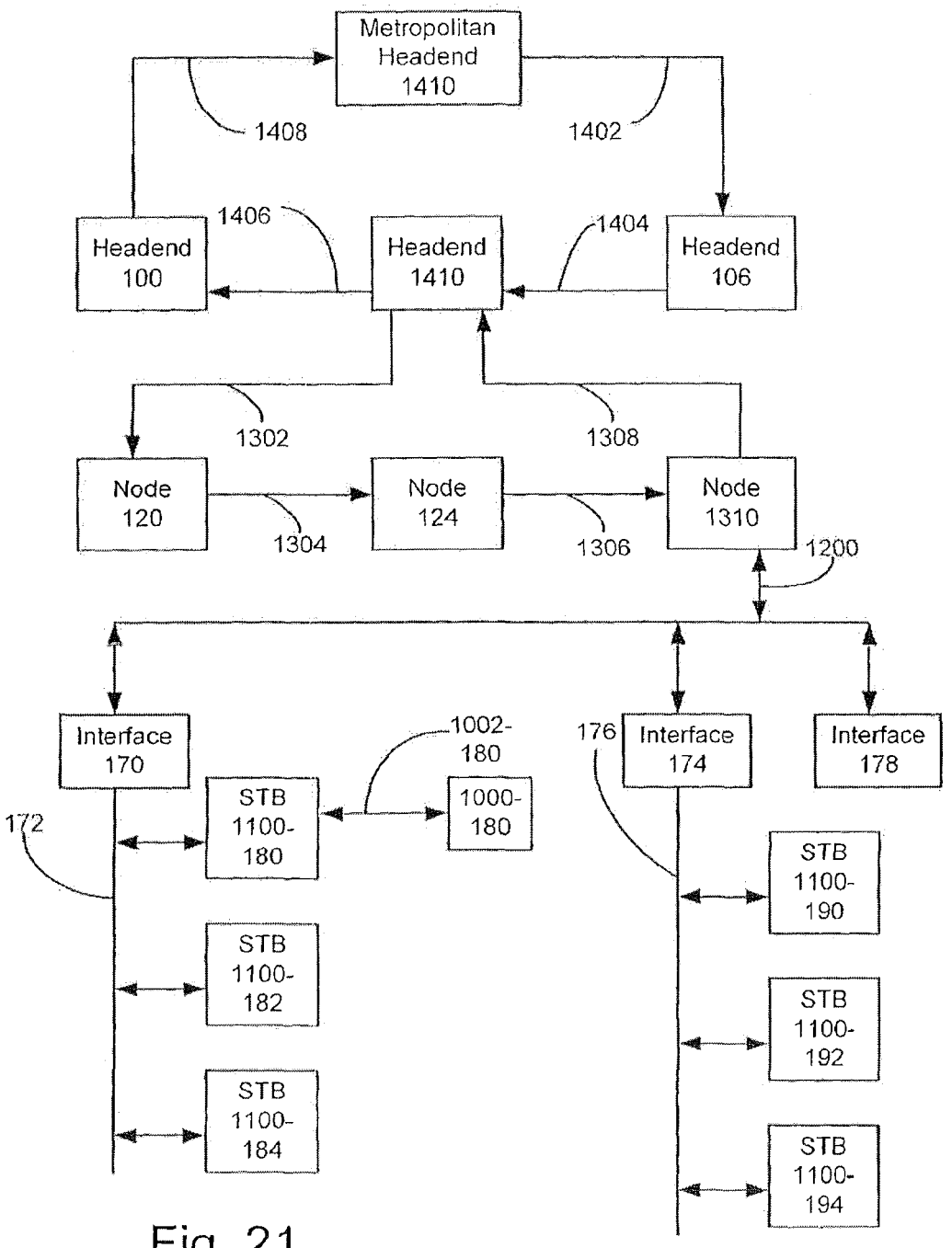

Metropolitan Headend 1410
1408
1402
Headend 100
1406
Headend 1410
1404
Headend 106
1302
1308
Node 120
Node 124
Node 1310
1304
1306
1200
Interface 170
1002-180
176
Interface 174
Interface 178
172
STB 1100-180
1000-180
STB 1100-190
STB 1100-182
STB 1100-192
STB 1100-184
STB 1100-194

Fig. 21

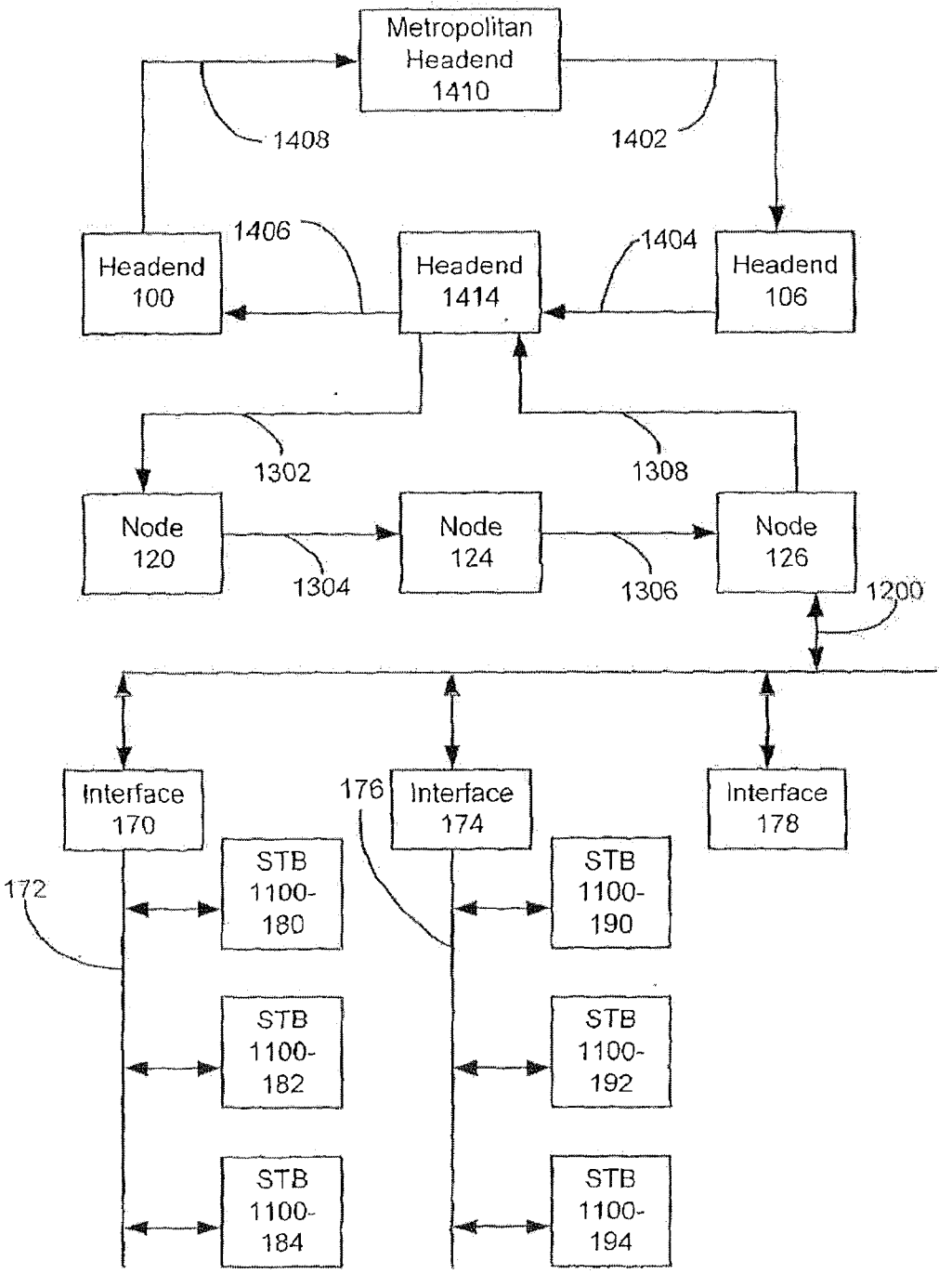

Metropolitan Headend 1410
1408
1402
1406
1404
Headend 100
Headend 1414
Headend 106
1302
1308
Node 120
Node 124
Node 126
1304
1306
1200
Interface 170
176
Interface 174
Interface 178
172
STB 1100-180
STB 1100-190
STB 1100-182
STB 1100-192
STB 1100-184
STB 1100-194

Fig. 22

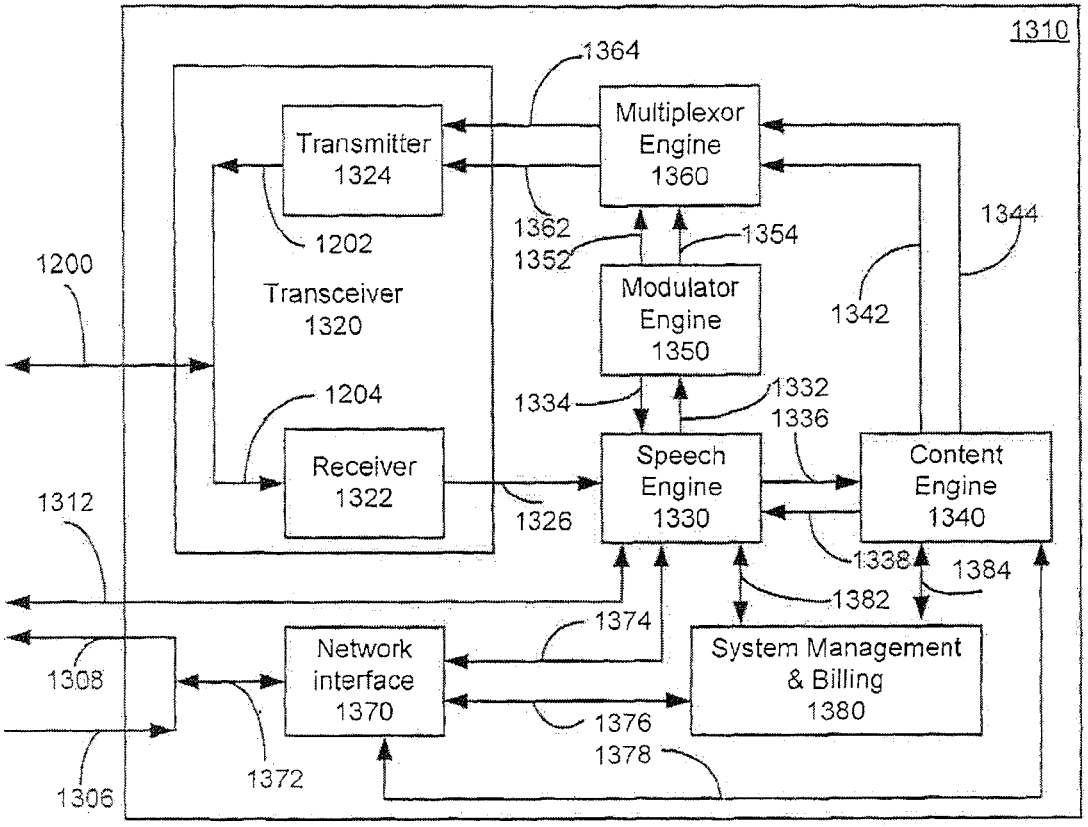

1310
1364
Multiplexor Engine 1360
Transmitter 1324
1202
1362
1352
1354
1344
1200
Transceiver 1320
Modulator Engine 1350
1342
1204
1334
1332
1336
Receiver 1322
Speech Engine 1330
Content Engine 1340
1312
1326
1338
1384
1382
1374
Network interface 1370
System Management & Billing 1380
1308
1376
1378
1372
1306

Fig. 23

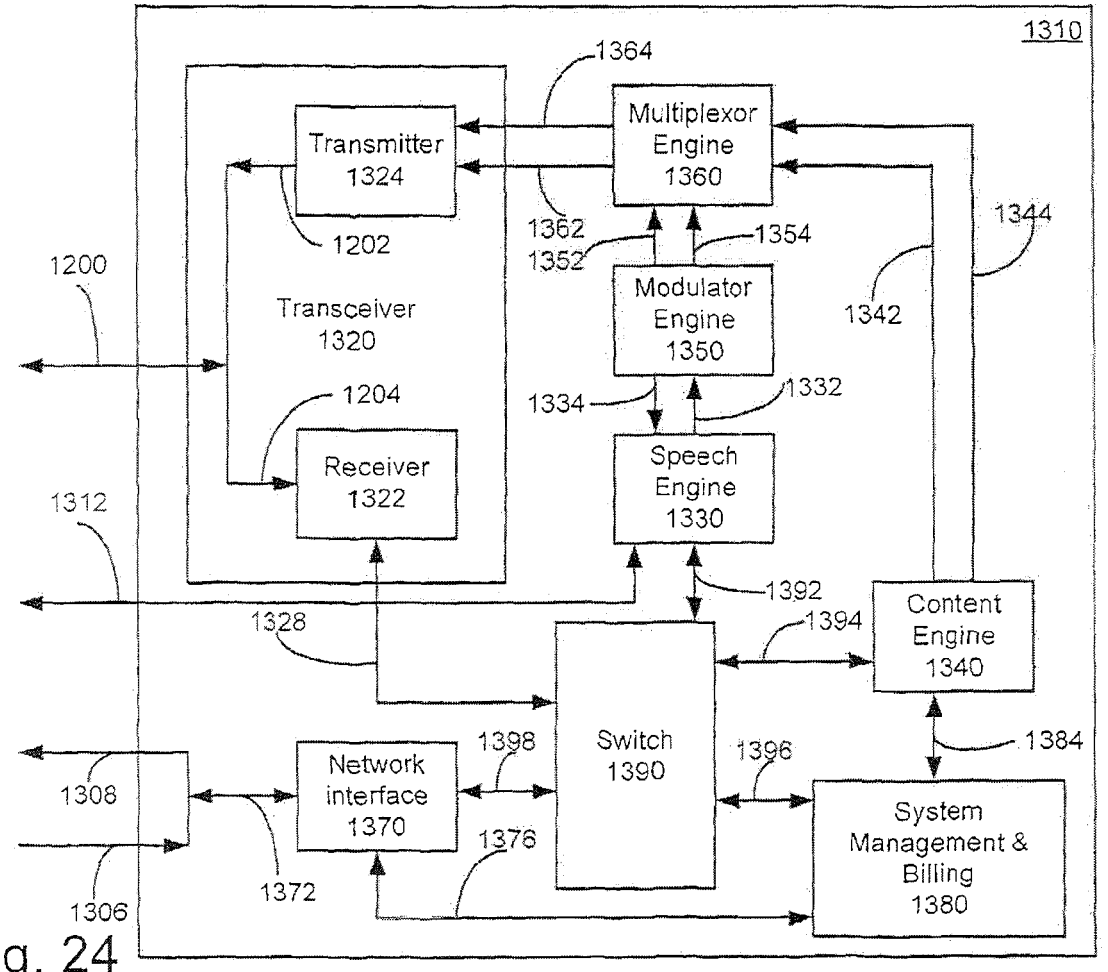

1310
1364
Multiplexor Engine 1360
Transmitter 1324
1202
1362
1352
1354
1344
1342
1200
Transceiver 1320
Modulator Engine 1350
1204
1334
1332
Receiver 1322
Speech Engine 1330
1312
1392
1328
1394
Content Engine 1340
Switch 1390
1398
1396
1384
1308
Network interface 1370
System Management & Billing 1380
1376
1372
1306

Fig. 24

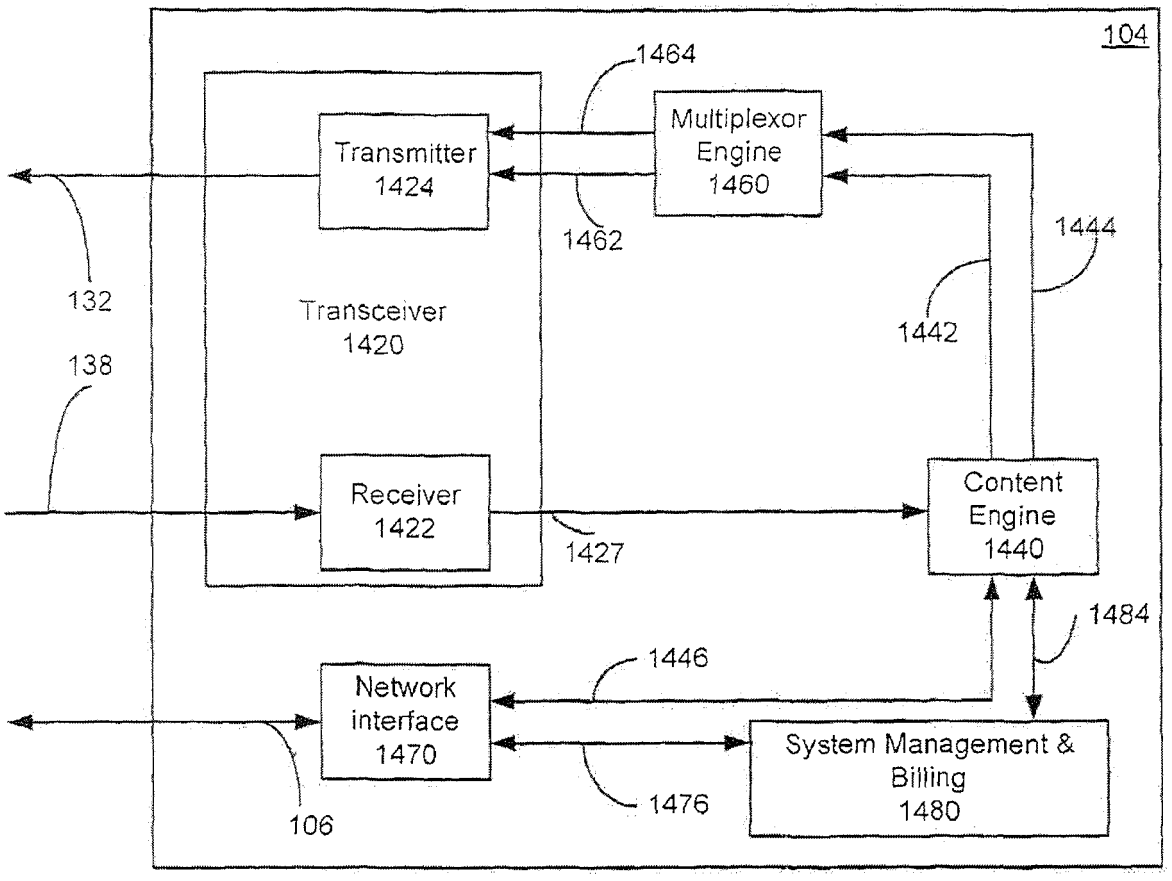


Fig. 25 Prior Art

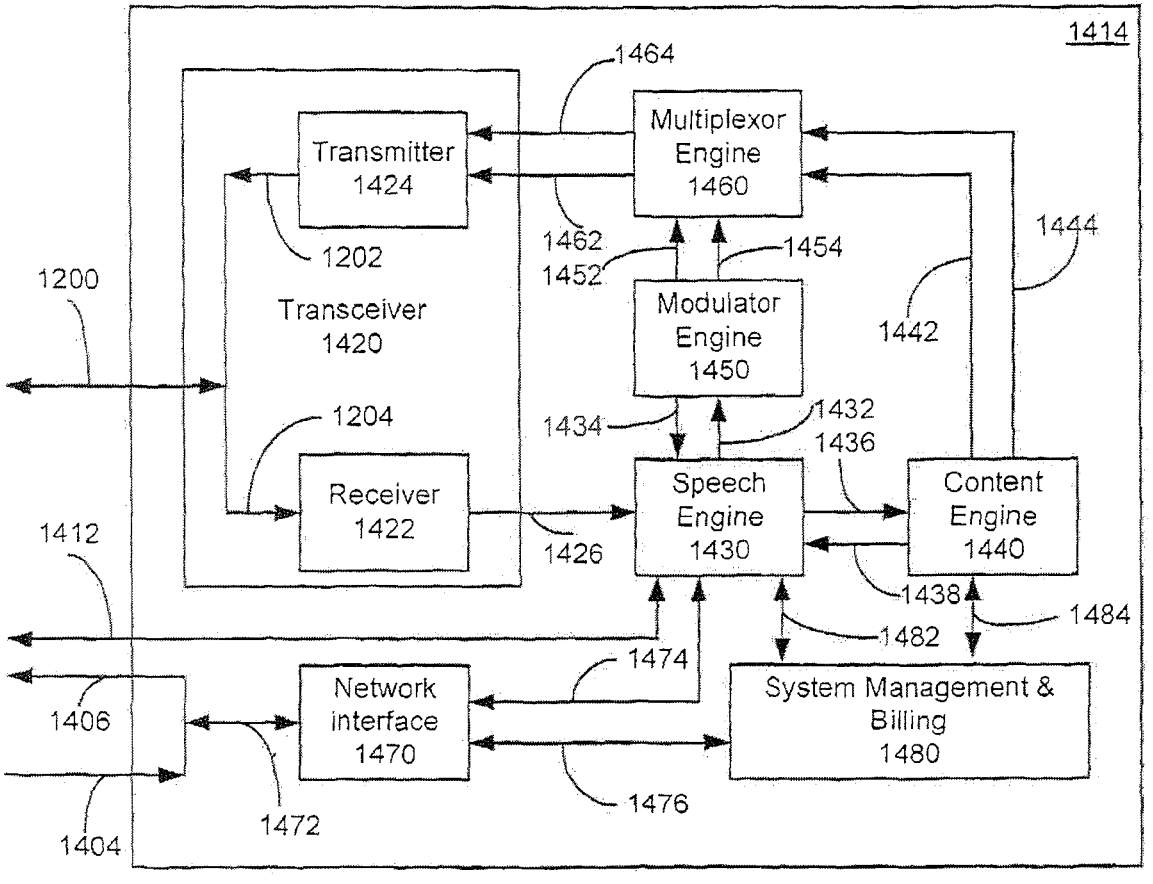

1414
1464
Multiplexor Engine 1460
Transmitter 1424
1202
1462
1452
1454
1444
1200
Transceiver 1420
Modulator Engine 1450
1442
1204
1434
1432
1436
Receiver 1422
Speech Engine 1430
Content Engine 1440
1412
1426
1438
1484
1482
1474
1406
Network interface 1470
System Management & Billing 1480
1476
1472
1404

Fig. 26

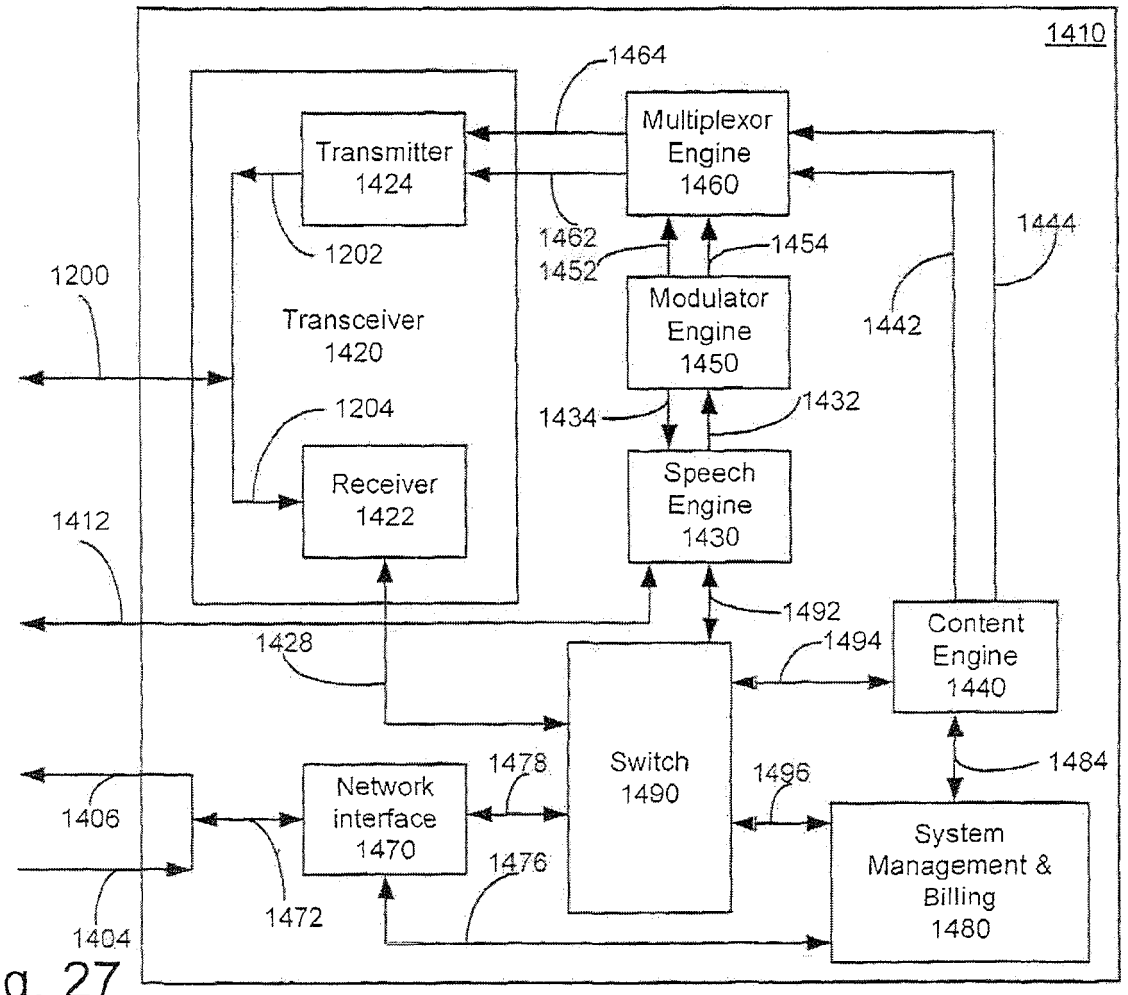

1410
1464
Multiplexor Engine 1460
Transmitter 1424
1202
1462
1452
1454
1444
1442
1200
Transceiver 1420
Modulator Engine 1450
1204
1434
1432
Receiver 1422
Speech Engine 1430
1412
1492
Content Engine 1440
1428
1494
Switch 1490
1478
1496
1484
1406
Network interface 1470
System Management & Billing 1480
1476
1472
1404

Fig. 27

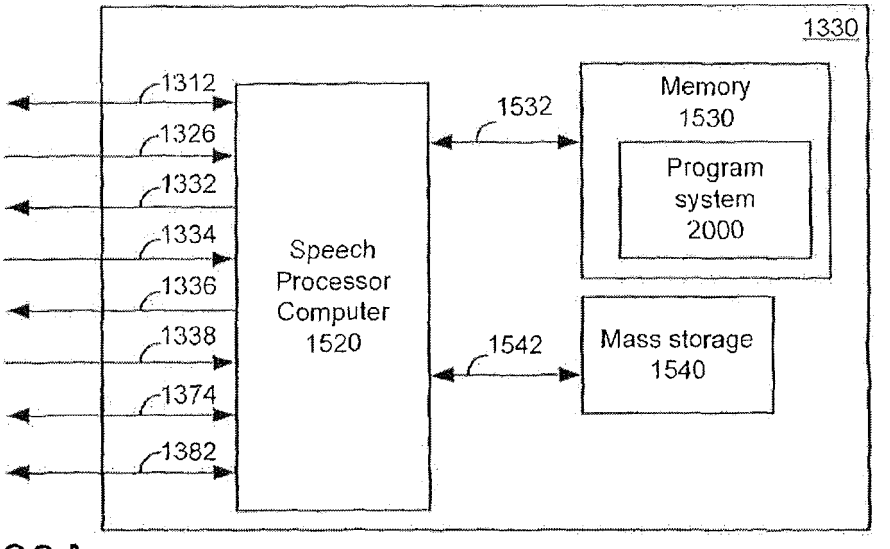


Fig. 28A

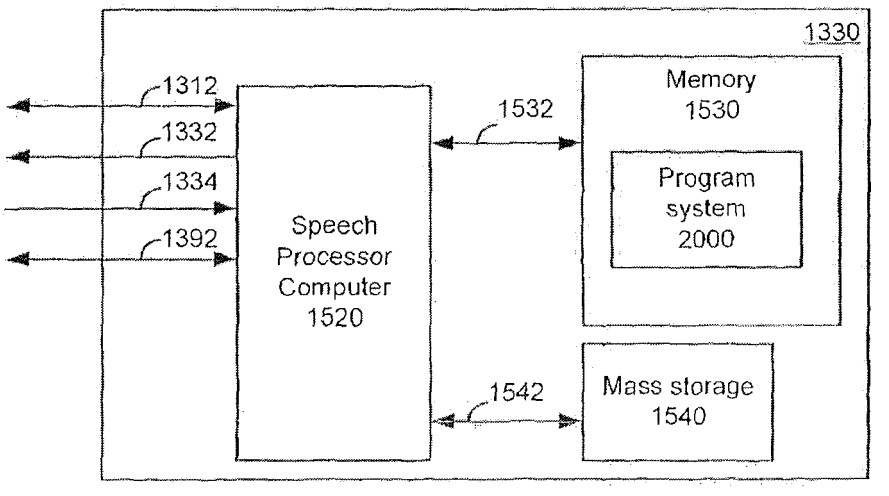


Fig. 28B

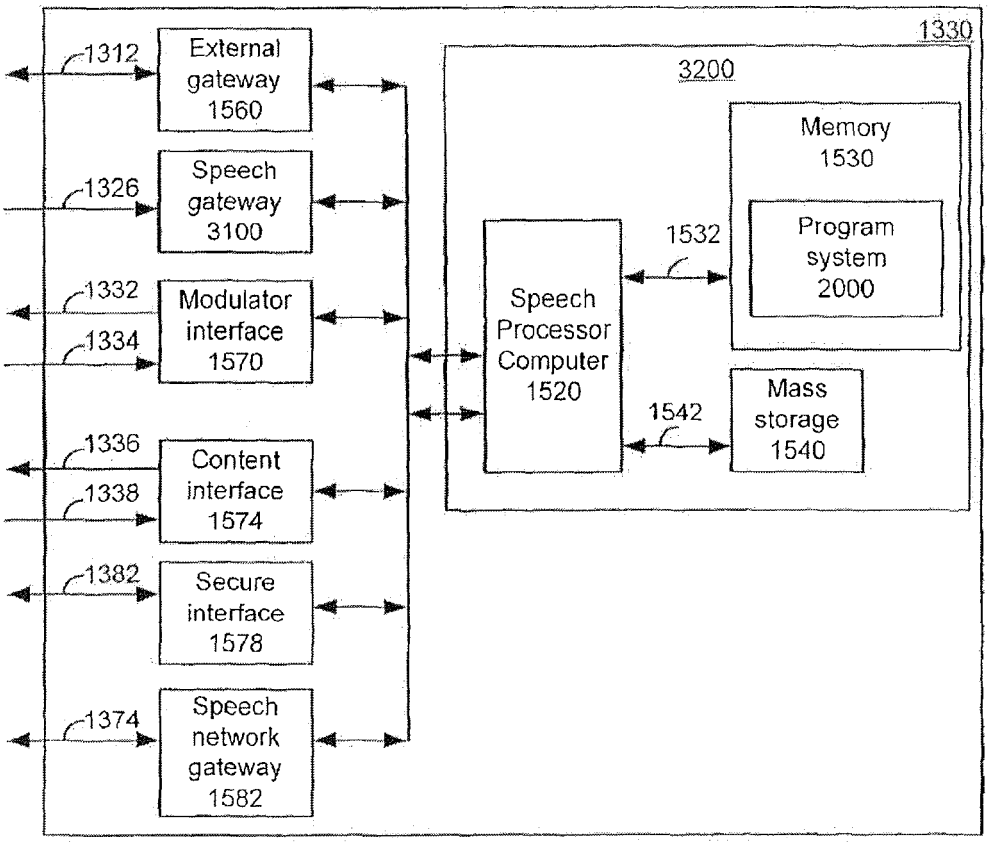

1330
3200
1312
External gateway 1560
1326
Speech gateway 3100
1332
1334
Modulator interface 1570
1336
1338
Content interface 1574
1382
Secure interface 1578
1374
Speech network gateway 1582
Speech Processor Computer 1520
1532
Memory 1530
Program system 2000
1542
Mass storage 1540

Fig. 29

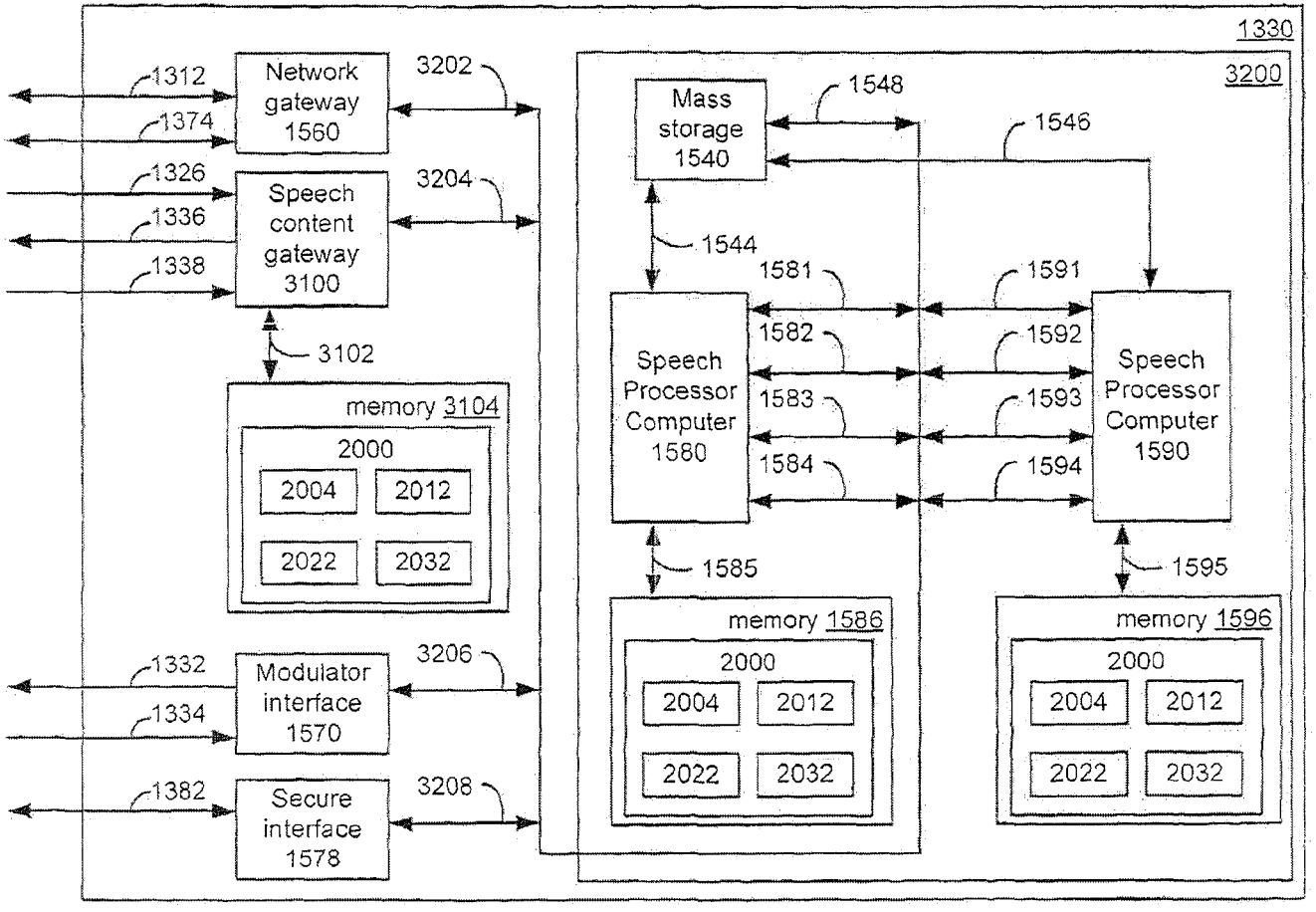

1330
3200
1312
1374
Network gateway 1560
3202
1326
1336
1338
Speech content gateway 3100
3204
3102
memory 3104
2000
2004
2012
2022
2032
1332
1334
Modulator interface 1570
3206
1382
Secure interface 1578
3208
Mass storage 1540
1548
1546
1544
1581
1582
1583
1584
Speech Processor Computer 1580
1585
memory 1586
2000
2004
2012
2022
2032
1591
1592
1593
1594
Speech Processor Computer 1590
1595
memory 1596
2000
2004
2012
2022
2032

Fig. 30

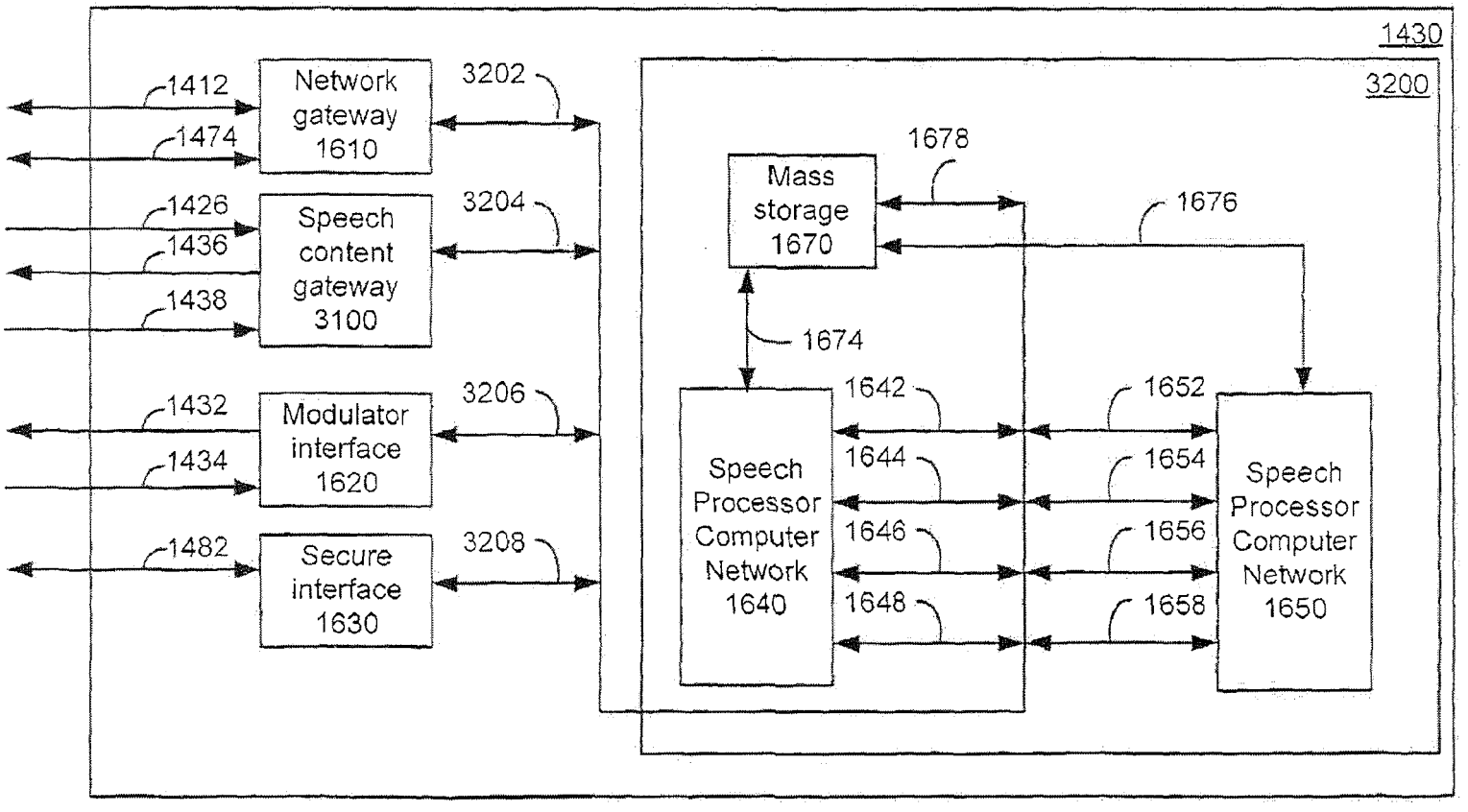

1430
3200
1412
1474
Network gateway 1610
3202
1426
1436
1438
Speech content gateway 3100
3204
1432
1434
Modulator interface 1620
3206
1482
Secure interface 1630
3208
1678
Mass storage 1670
1676
1674
1642
1644
1646
1648
Speech Processor Computer Network 1640
1652
1654
1656
1658
Speech Processor Computer Network 1650

Fig. 31

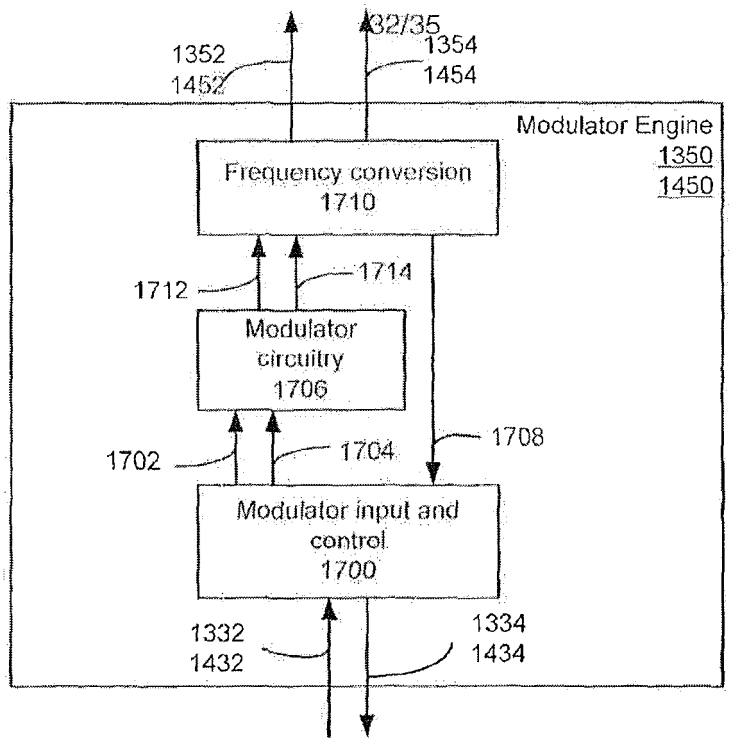


Fig. 32A

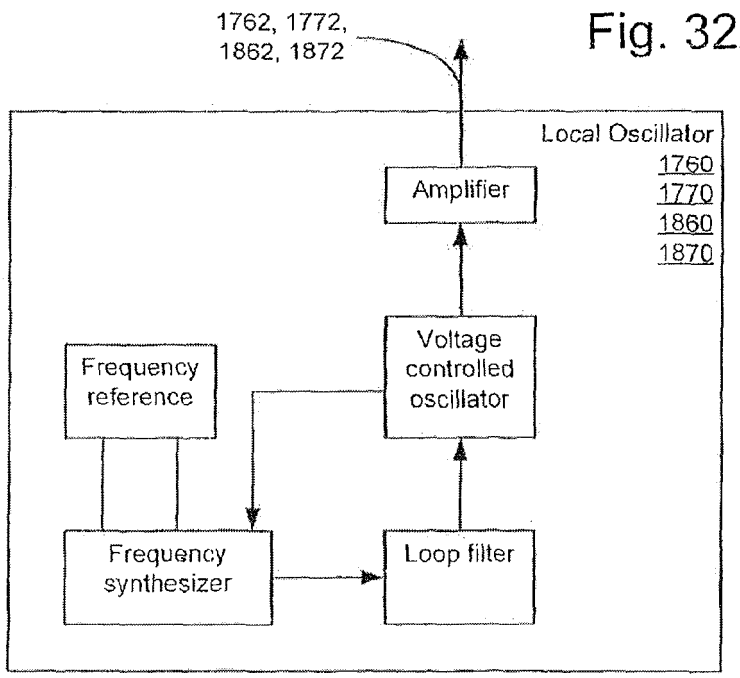


Fig. 32B Prior Art

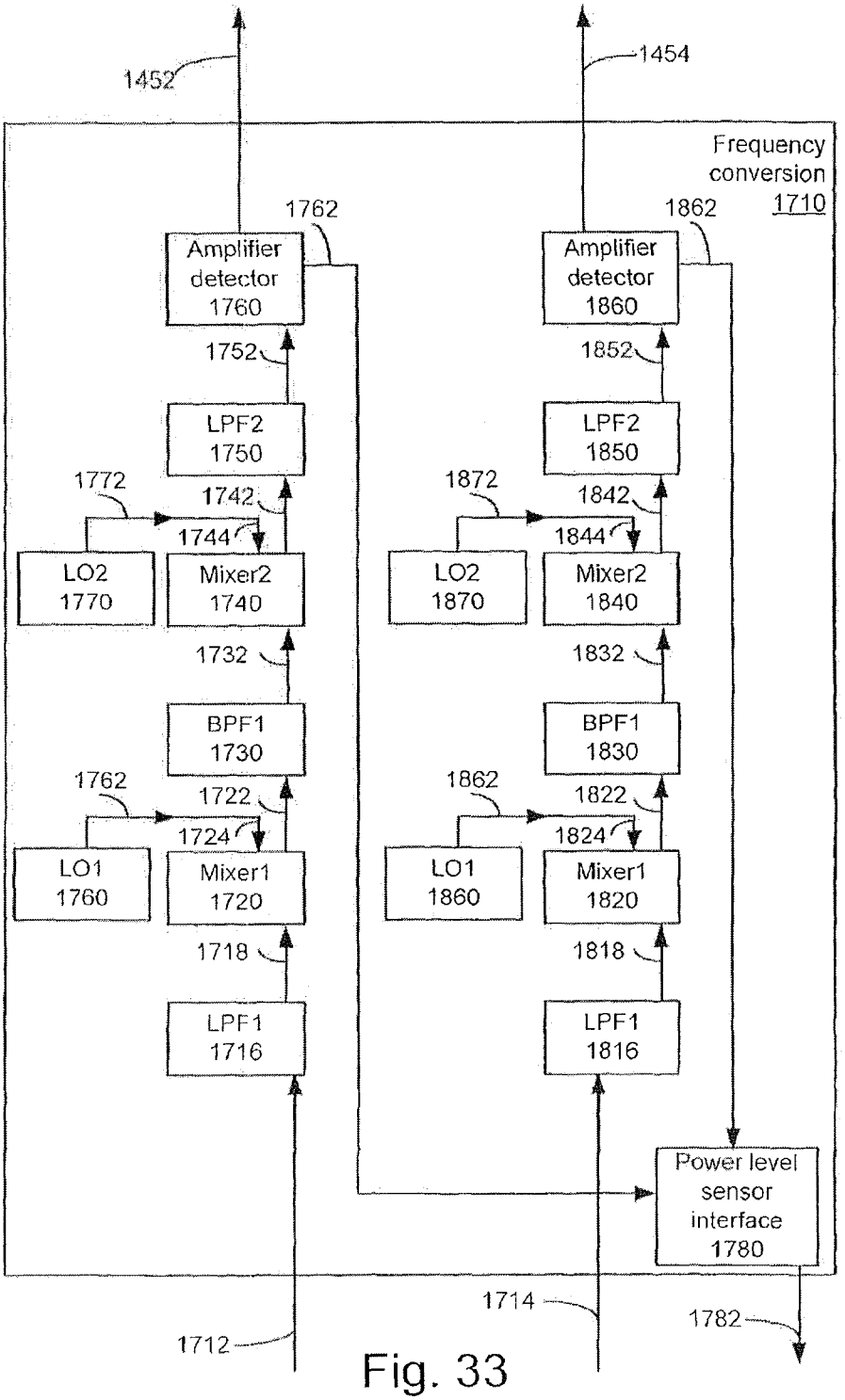

1452
1454
Frequency conversion 1710
1762
1862
Amplifier detector 1760
Amplifier detector 1860
1752
1852
LPF2 1750
LPF2 1850
1772
1742
1744
1872
1842
1844
LO2 1770
Mixer2 1740
LO2 1870
Mixer2 1840
1732
1832
BPF1 1730
BPF1 1830
1762
1722
1724
1862
1822
1824
LO1 1760
Mixer1 1720
LO1 1860
Mixer1 1820
1718
1818
LPF1 1716
LPF1 1816
Power level sensor interface 1780
1712
1714
1782

Fig. 33

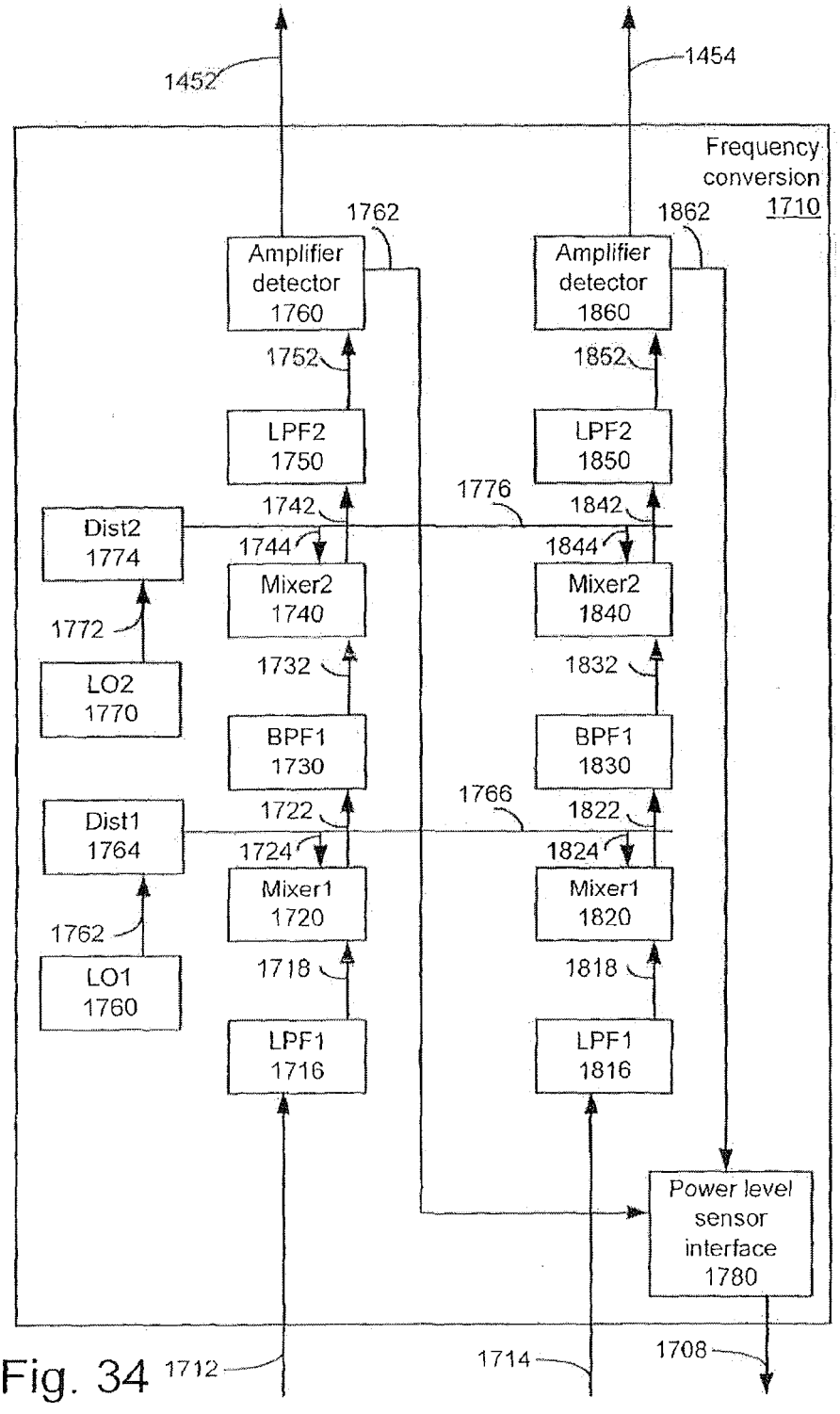

1452
1454
Frequency conversion 1710
1762
1862
Amplifier detector 1760
Amplifier detector 1860
1752
1852
LPF2 1750
LPF2 1850
1776
1742
1842
Dist2 1774
1744
1844
Mixer2 1740
Mixer2 1840
1772
LO2 1770
1732
1832
BPF1 1730
BPF1 1830
1766
1722
1822
Dist1 1764
1724
1824
Mixer1 1720
Mixer1 1820
1762
LO1 1760
1718
1818
LPF1 1716
LPF1 1816
Power level sensor interface 1780
1712
1714
1708

Fig. 34

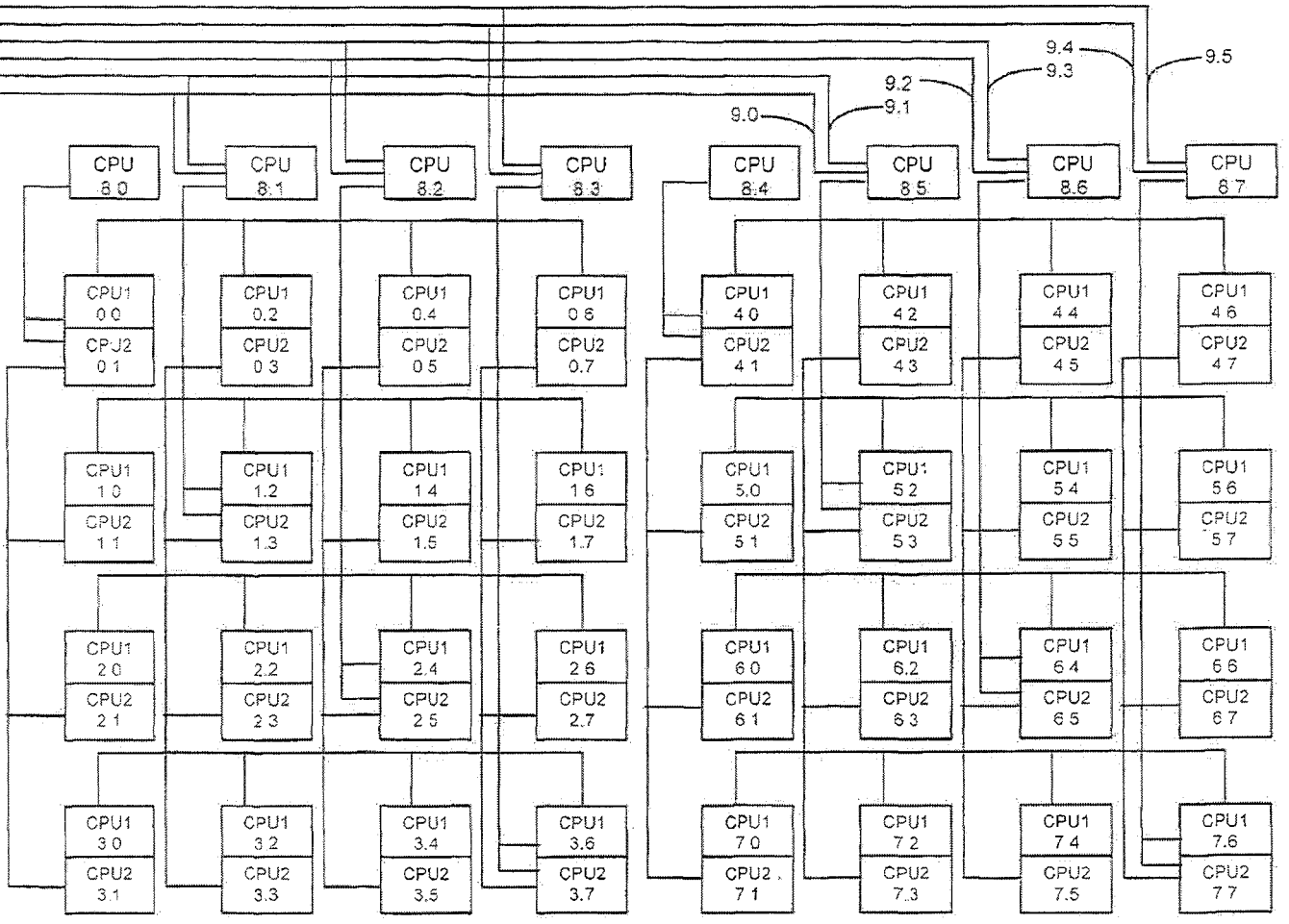

9.0
9.1
9.2
9.3
9.4
9.5
CPU 8.0
CPU 8.1
CPU 8.2
CPU 8.3
CPU 8.4
CPU 8.5
CPU 8.6
CPU 8.7
CPU1 0.0
CPU2 0.1
CPU1 0.2
CPU2 0.3
CPU1 0.4
CPU2 0.5
CPU1 0.6
CPU2 0.7
CPU1 4.0
CPU2 4.1
CPU1 4.2
CPU2 4.3
CPU1 4.4
CPU2 4.5
CPU1 4.6
CPU2 4.7
CPU1 1.0
CPU2 1.1
CPU1 1.2
CPU2 1.3
CPU1 1.4
CPU2 1.5
CPU1 1.6
CPU2 1.7
CPU1 5.0
CPU2 5.1
CPU1 5.2
CPU2 5.3
CPU1 5.4
CPU2 5.5
CPU1 5.6
CPU2 5.7
CPU1 2.0
CPU2 2.1
CPU1 2.2
CPU2 2.3
CPU1 2.4
CPU2 2.5
CPU1 2.6
CPU2 2.7
CPU1 6.0
CPU2 6.1
CPU1 6.2
CPU2 6.3
CPU1 6.4
CPU2 6.5
CPU1 6.6
CPU2 6.7
CPU1 3.0
CPU2 3.1
CPU1 3.2
CPU2 3.3
CPU1 3.4
CPU2 3.5
CPU1 3.6
CPU2 3.7
CPU1 7.0
CPU2 7.1
CPU1 7.2
CPU2 7.3
CPU1 7.4
CPU2 7.5
CPU1 7.6
CPU2 7.7

Fig. 35

SYSTEM AND METHOD OF VOICE RECOGNITION NEAR A WIRELINE NODE OF A NETWORK SUPPORTING CABLE TELEVISION AND/OR VIDEO DELIVERY

This application claims priority from Provisional Patent Application No. 60/210,440, entitled Method and Apparatus for Centralized Voice-Driven Natural Language Processing, filed 8 Jun. 2000.

This application is a continuation in part from application Ser. No. 09/679,115, entitled System and Method of a Multi-dimensional Plex Communication Network, filed 4 Oct. 2000;

Application Ser. No. 09/664,874, entitled Increased Bandwidth In Aloha-Based Frequency-Hopping Transmission Systems, filed 19 Sep. 2000 now Abandoned;

Application Ser. No. 09/661,486, entitled N-way Demultiplexor, filed 14 Sep. 2000; and

Application Ser. No. 09/708,315, entitled Common Carrier, Multiple Output RF Upconverter, filed 7 Nov. 2000 now U.S. Pat. No. 6,480,703.

BACKGROUND OF THE INVENTION

1. Technical Field

This invention relates to speech recognition performed near a wireline node of a network supporting cable television and/or video delivery.

2. Background Art

Currently, speech operated functions using the latest speech recognition technologies are limited to a handful of applications, such as toys, appliances, some computers, speech dictation, cellular phones, and speech control of one's home. Most of these applications use speech recognition technology running on a computer or speech recognition chip technology. These speech recognition systems typically offer only a limited number of commands and the recognition efficiency is only fair and often require speech training.

There have been numerous patents issued regarding speech recognition. Many apply in a telephone context or other dial-up context such as an Automated Teller machine (ATM), including the following: Rabin, *Voice command control and verification system*, U.S. Pat. No. 6,081,782, issued Jun. 27, 2000, Basore, et al, *Voice activated device and method for providing access ro remotely retrieved data*, U.S. Pat. No. 5,752,232, issued May 12, 1998, and Kowalkowski, et al, *Voice-control integrated field support data communications system for maintenance, repair and emergency services*, U.S. Pat. No. 5,924,069, issued Jul. 13, 1999.

There is, however, another class of speech recognition technology referred to as natural language, which requires state of the art processing software and hundreds of megabytes of RAM to support. Natural language recognition is currently being used in high end systems, such as billing applications for utility companies and the New York Stock Exchange, because of its ability to recognize spoken words from any speech. Some natural language systems claim to be totally user independent and are also capable of recognizing speech in several different languages.

However, the problems of speech recognition at a centralized wireline node in a network supporting video delivery or cable television delivery have not been addressed by such prior art. For the purposes of the discussion herein, a centralized wireline node refers to a network node providing

video or cable television delivery to multiple users using a wireline physical transport between those users at the node.

FIG. 1 depicts a typical network as found in a cable television and/or video delivery network employing a Hybrid Fiber-Coaxial (HFC) wiring scheme as disclosed in the prior art.

Each user site contains a Set Top Box, such as STB **180**, coupling to the network through a coaxial cable **172**, which interfaces **170** to a collective coaxial cable **160** which couples to a Node **126**. The interface **170** may include bi-directional signal amplification and possibly further include the filtering and/or frequency shifting of these signals.

The Node **126** is hierarchically coupled **128** to a Headend **104**, which in most cable television networks serves as the source of television programming and other signaling. The signals are sent through the Node **126** and couplings **160-170-172** to provide the STB **180** and others, with the television signaling. In certain large towns and cities, there may be a further hierarchical layer including a Metropolitan Headend **10** coupled **106** to Headend **104**. These higher layers of the network use fiber optics for the physical transport of couplings **102**, **106** and **108**, as well as for **122**, **126** and **128**.

The couplings between STB **180** and Node **126** support bi-directional communication. The couplings between STB **180**, Node **126** and Headend **104** may also support bi-directional communication. Such bi-directional communication allows the STB **180** to receive multiple television channels. This bi-directional communication allows STB **180** to signal at least limited information to the Node **126** and/or the Headend **104**. Such information in either case may support management of Pay-per-View and other services.

User site accounting information usually resides at the highest level of the network, which tends to be either the Headend **104** or Metropolitan Headend **10**.

In cable systems, several downstream data channels that send channel and synchronization information are often transmitted in a previously reserved band of frequencies. They are typically assigned for re-broadcasting FM channels over cable in the United States. Currently, most cable systems reserve some of the 88 to 108 MHz FM spectrum for set-top data transmission. The unused portion of that spectrum are left for barker channels or for additional video channels. The Open Cable Standard requires that the 70 to 130 MHz band be available for what's called Out-of-Band or (OOB) or Downstream transmission.

Most current cable systems use the popular HFC architecture so that the downstream video signals, digital or analog, are sent from the Headend to hubs or nodes via fiberoptic cable. At the receiving side of the node, the optical signal from the fiber gets converted to an electrical signal containing all of the analog, digital video RF carriers and program/service information. This signal, in turn, is amplified and distributed via coaxial cable to the appropriate subscribers connected to the node.

A major design objective for existing cable television set-top boxes was efficient downstream information delivery, i.e. from cable plant to subscriber. Provision for upstream data transmission, i.e. from subscriber to cable plant, is much more restrictive, supporting only limited bandwidth. As new classes of interactive services become available, efficient use of upstream transmission bandwidth grows in importance. For example, if it is necessary to pass

information from the subscriber to the cable headend (also known as the headend), sufficient upstream bandwidth must be made available.

One of the most popular digital set-top boxes, the General Instruments (now Motorola) DCT-2000, is a useful example. When this box was first deployed, upstream transmissions were restricted to user pay-per-view requests, and other simple, infrequent transmissions. As a consequence, the transmission format used for upstream transmissions was not required to be very efficient, and in fact, is not.

In this set-top box, the transmission hardware is capable of selecting twenty different 256K bps channels, each of which uses QPSK transmission coding. While the hardware is capable of frequency-hopping to avoid channels which are subject to interference, the scheme used is fairly static, with typical deployments only using two active upstream communications channels. This leads to an aggregate bandwidth of only 512K bps per cluster of set-top boxes converging in the network to a node, in cable television terms. The cable node typically supports between 500 and 2000 subscribers.

Upstream signals in the 5 to 40 MHz band from each subscriber connected to the node are collected, combined, and then sent to the Headend via either the same fiber used for the downstream video carriers, or a separate fiber.

Furthermore, the transmission control protocol used, referred to as Aloha, is one where an individual set-top box immediately transmits any pending request to the headend, without regard to whether or not the transmission channel is already in use. This transmission is repeated at regular intervals until the box receives an acknowledgement command from the headend, indicating successful receipt of the transmission.

This transmission control protocol is quite inefficient due to the number of collisions which ensue, e.g. simultaneous transmissions from different set-top boxes which interfere with one another, forcing all of the transmitters to repeat their transmissions again. This leads to typical channel utilization on the order of just 30%. As a consequence, the total bandwidth available for upstream transmission per node is only about 30% of 512K bps=~137K bps, on average.

Downstream control data transmission typically occurs in a separate frequency band from the upstream channels.

Typically, HFC networks employ an optical fiber from a central office, or Headend, to a neighborhood node. The fiber has forward and reverse transmission capability, which can alternatively be accommodated on separate fibers. Wavelength Division Multiplexing (WDM) can be used to implement both on a single fiber. At the node, coaxial cable connects the users through a shared frequency division multiplexing (FDM) scheme with contention resolution protocols used to manage upstream data flows.

Such communication schemes having both forward and backward paths, and which may or may not involve a user, are referred to as loops herein. An example of a loop is the communication between Headend **104** and Node **126**. Communication schemes having both forward and backward paths to multiple users are referred to as local loops. An example of a local loop is the communication between Node **126** and user site STBs **180**, **182** and **184**. Note that a loop may be constituted out of optical fiber or out of coaxial cable.

Hybrid-Fiber-Copper (HFCop) networks work in much the same manner, but substitute copper wire(s), often in twisted pairs, for coaxial cable. In such networks a local loop may further be constituted out of optical fiber, coaxial cable or twisted pairs.

Another alternative local loop configuration is commonly known as Switched Digital Video. It is a form of HFC coupling the fiber through a node to each user site with a distinct point-to-point coaxial cable. The node interfaces the user site coaxial cables with the optical fiber through a switch. The switch typically contains a network management unit which manages the switch, connecting the bandwidth service provider with multiple homes, today often in the range of five to 40 homes per switch.

The Synchronous Optical NETwork (SONET) scheme is also applied in the creation of high-speed networks for homes and businesses. This and similar communication schemes may be employed to deliver video streams to user sites.

FIG. **2** depicts a typical residential broadband network using local loop wiring of the network, as disclosed in the prior art.

As in FIG. **1**, each user site contains a Set Top Box, such as STB **180**, coupled to the network through a coaxial cable **172** which interfaces **170** to a collective coaxial cable **160** which is coupled to Node **126**. Interface **170** may include bi-directional signal amplification, and possibly further include the filtering and/or frequency shifting of these signals.

As in FIG. **1**, the couplings between STB **180** and Node **126** support bi-directional communication allowing the STB **180** to receive multiple television channels and allowing STB **180** to signal at least limited information to the Node **126**, which may well include management of Pay-per-View and other services. The couplings between STB **180**, Node **126** and Headend **104** may also support bi-directional communication allowing the STB **180** to receive multiple television channels and allowing STB **180** to signal at least limited information to the Headend **104**, which may well include management of Pay-per-View and other services.

FIG. **2** shows a loop coupling Headend **104** through coupling **130** to Node **120** through coupling **132** to Node **124** through coupling **134** to Node **126** which in turn couples **136** to Headend **104** forming the loop.

The hierarchical coupling of Node **126** with Headend **104** is carried out along distinct paths through this loop. Communication from Headend **104** to Node **126** follows a path **130-132-134**. Communication from Node **126** to Headend **104** follows the path **136**. The specific wiring schemes are dominated by the choice of physical transport, communication protocols and network level management. The description just given for FIG. **2** is provided as a simplified discussion of the basics of how high speed residential broadband networks incorporate loops and local loops supporting network level hierarchies.

There has been extensive research into the mechanics of speech recognition. The progress has been sufficient to allow speech trading by stock brokers using their desk top computers.

While these innovations have been substantial, they do not resolve several central questions of great importance to cable television, video delivery systems, and commerce. There is no present system providing speech recognition to a collection of users over a cable television network. There is no present system providing user identification based upon that speech recognition over a network that supports cable television and/or video delivery. There is no present system sufficient for real-time auctions and contracting to be conducted over a cable television and/or video delivery network, based on user identification through speech recognition.

SUMMARY OF THE INVENTION

An embodiment of the invention provides speech recognition services to a collection of users over a network that supports cable television and/or video delivery. User identification based upon speech recognition is provided over a cable television and/or video delivery network. User identified speech contracting is provided over a cable television and/or video delivery network having sufficient bandwidth for real-time auctions and contracting.

The invention comprises a multi-user control system for audio visual devices that incorporates a speech recognition system that is centrally located in or near a wireline node, and which may include a Cable Television (CATV) Headend. The speech recognition system may also be centrally located in or near a server farm a web-site hosting facility, or a network gateway.

In these embodiments of the invention, spoken commands from a cable subscriber are recognized and then acted upon to control the delivery of entertainment and information services, such as Video On Demand, Pay Per View, Channel control, on-line shopping, and the Internet. This system is unique in that the speech command which originates at the user site, often the home of the subscriber, is sent upstream via the return path (often five to 40 MHz) in the cable system to a central speech recognition and identification engine. The speech recognition and identification engine described herein is capable of processing thousands of speech commands simultaneously and offering a low latency entertainment, information, and shopping experience to the user or subscriber.

The system is capable of overlaying text on the subscriber's TV screen immediately after a word is recognized by the system as a verification of correct or incorrect recognition, thereby providing instant visual feedback and opportunity for acceptance or correction of speech messages.

The system can recognize and process speech so that the key words of spoken commands are recognized and displayed. This may be applied in navigation mode. in search context, or in other contexts and modes.

The system responds to a command with a visual indication of the spoken request. This visual feedback indicates recognition of key words may and be in the form of written text or icons.

The system may mask delays in upstream transmission or speech recognition. Upon depressing an optional talk button on the remote, a digital address may be appended to the beginning of the digital speech packets to be processed. The address tells the system not only the user site, but it also provides a mechanism for the system to begin generating masking screens or icons. Data are sent to the central location when the button on the microphone is depressed, alerting the system as to the user site and a potential input. This function allows the system to generate an icon or overlay to respond to the subscriber quickly. This function also supports site specific dictionaries, as well as data references to be loaded for speech recognition or user recognition.

At least two operations are performed at a server-center located at a central location: upstream recognition of speech commands and performing speech command protocol(s).

Low latency visual prompts are provided to support fast and accurate speech navigation. Thus, when a command is spoken and recognized the system returns the key word and optionally generates a list of visual prompts that guides the subscriber through the next navigation step. By using prompts, the system incorporates optimum recognition of the prompt words thus increasing the recognition accuracy and, at the same time, increasing satisfaction for user participation.

In this context of the invention, adaptive speech recognition is based on optimized word recognition by creating a subset of probable matches based on knowing what is in an entertainment database or the words on a Web page. This supports learning the user's habitual speech patterns.

Secure speech transactions provide a method of speech based contracting across a multimedia distribution system. Speech based contracting as used herein, refers to a process of generating and/or executing a contract, in which at least one step of that process is based upon the invention receiving, recognizing, and witnessing an identified user's speech.

Speech based contracting includes, but is not limited to, the following: a first identified user making an offer; a second identified user accepting an offer, which may or may not be acoustically presented; as well as the second identified user specifying acceptance particulars, such as payment arrangements and the number of units.

Speech based contacting also includes, but is not limited to, acknowledging receipt of goods or services of a tangible, and/or intangible nature, possibly involving real property, personal property and/or intellectual property, exercising options of the contract, as well as terminating a pre-existing contract. The acknowledgement of receipt may include, but is not limited to, a declaration of the condition of goods upon receipt or upon subsequent testing, which may include an estimate of damage.

Speech based contracting may also include at least one of the following: the second identified user making a counter-offer to the first identified user based upon the offer; and the first identified user responding to the counter offer. The response may include accepting the counter-offer, or making a second counter-offer to the second identified user.

Speech based contracting may also include the second identified user inviting offers. The invitation may or may not specify the first identified user and/or performance constraints such as time of delivery and/or the duration of the terms and/or optional terms which may be exercised after the contract has been accepted.

Speech based contracting may also include an identified user exercising one or more optional terms of a pre-existing contract. The exercise of the optional terms may further specify one or more amounts, delivery times, ranges of time and/or place during/over which a service may be performed or commodity delivered in accordance with the optional terms of the pre-existing contract.

The offer may originate at the offeror-user site, possibly involving speech, when it is sent to the central location and recognized at the central location, recorded, distributed, and presented to potential offeree user sites. The offer may be sent to the central location to be recorded and distributed to potential offeree user sites. The offeree may signal acceptance of the offer verbally at the offeree user site where the offeree is recognized. The system transmits the verbal acceptance to the central location, where it is recognized, recorded and then transmitted to the offeror.

Using the user site address accelerates speaker identification, increases reliability, enhances security and reduces latency in identifying a speaker. Using the user site address and user specific speech data references further accelerates speaker identification, increases reliability, enhances security, and reduces latency in identifying a speaker.

The invention supports automatic generation of serial-multiplexed video output streams, without additional video boards or multiplexers. The centrally located speech recog-

nition system employs extremely fast, efficient arrays of microprocessors, many of which may possess a frame buffer in locally accessible memory. Each microprocessor translates the frame buffer into an MPEG stream. Several MPEG streams are merged within a single microprocessor to form a multi-media stream for distribution to subscribers.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1** depicts a typical network hierarchy as found in a cable television or video delivery network employing a Hybrid Fiber-Coaxial (HFC) wiring scheme as disclosed in the prior art;

FIG. **2** depicts a typical residential broadband network using local loop wiring of the network as disclosed in the prior art;

FIG. **3** depicts a remote control unit **1000** coupled **1002** to set-top apparatus **1100**, communicating via a two-stage wireline communications system containing a wireline physical transport **1200** through a distributor node **1300**, and through a high speed physical transport **1400**, possessing various delivery points **1510** and entry points **1512**–**1518** to a tightly coupled server farm **3000**, with one or more gateways **3100**, and one or more tightly coupled server arrays **3200**, in accordance the invention;

FIG. **4** depicts a coupled server array **3200** of FIG. **3**;

FIG. **5** depicts a gateway **3100** of FIG. **3**;

FIG. **6** depicts the tightly coupled server farm **3000** of FIG. **3** implemented in a two-dimensional plex communication network with N=4 plex nodes in each of two orthogonal directions of the node array;

FIG. **7** depicts a gateway **3100** of FIG. **3** implemented in a two-dimensional plex communication network with N=4 plex nodes in each of two orthogonal directions of the node array;

FIG. **8** depicts a coupled server array **3200** of FIG. **3** implemented in a two-dimensional plex communication network with N=4 plex nodes in each of two orthogonal directions of the node array;

FIG. **9** depicts a simplified block diagram using an array of processors as shown as **3200** in FIG. **3**;

FIG. **10** depicts a flowchart of a method using a back channel from a multiplicity of user sites containing a multiplicity of identified speech channels presented to a speech processing system at a wireline node in a network supporting cable television delivery in accordance with the invention;

FIG. **11**A depicts a detail flowchart of operation **2012** of FIG. **10** further partitioning of the received back channel;

FIG. **11**B depicts a detail flowchart of operation **2022** of FIG. **10** further processing the multiplicity of the received identified speech channels;

FIG. **11**C depicts a detail flowchart of operation **2032** of FIG. **10** further responding to the identified speech content;

FIG. **12** depicts a detail flowchart of operation **2000** of FIG. **10** further performing the method using the back channel from multiple user sites;

FIG. **13**A depicts a detail flowchart of operation **2112** of FIG. **11**C further responding to the identified speech content;

FIG. **13**B depicts a detail flowchart of operation **2112** of FIG. **11**C further responding to the speech content;

FIG. **14** depicts a detail flowchart of operation **2112** of FIG. **11**C further responding to the identified speech content from the associated user site;

FIG. **15**A depicts a detail flowchart of operation **2112** of FIG. **11**C further responding to the identified speech content from the associated user site;

FIG. **15**B depicts a detail flowchart of operation **2252** of FIG. **15**A identifying the user;

FIG. **16** depicts a detail flowchart of operation **2112** of FIG. **11**C further responding to the identified speech content from the associated user site;

FIG. **17**A depicts the speech content response **2350** including current response menu **2352** and cumulative user site response **2354** in accordance with the invention;

FIG. **17**B depicts a detail flowchart of operation **2112** of FIG. **11**C further responding to the identified speech content from the associated user site;

FIG. **18**A depicts a detail flowchart of operation **2112** of FIG. **11**C further responding to the identified speech content from the associated user site;

FIG. **18**B depicts a detail flowchart of operation **2092** of FIG. **11**B further processing the multiplicity of the received speech channels;

FIG. **19**A depicts a simplified block diagram of a hand held remote **1000**, containing microphone **1060** and keypad **1020** supporting user input which is organized and processed by embedded controller **1050** for communication by wireless interface **1040** coupled **1002** to set-top apparatus **1100**, as shown in FIG. **3**;

FIG. **19**B depicts a simplified block diagram of set-top apparatus **1100** as shown in FIG. **3** showing coupling **1002** and first wireline physical transport **1200** further comprised of downlink coupling **1202** and uplink coupling **1204**;

FIG. **19**C further depicts set-top apparatus **1100** as shown in FIG. **19**B containing a set-top appliance **1120** coupled **1002** with hand held remote **1000** and coupled **1112** with set-top box **1120** possessing downlink coupling **1202** and uplink coupling **1204**;

FIG. **19**D further depicts set-top apparatus **1100** as shown in FIG. **19**B containing a set-top appliance **1120** coupled **1002** with hand held remote **1000** and possessing downlink coupling **1202** and uplink coupling **1204** as well as providing processed downlink coupling **1114** to set-top box **1110** and receiving initial uplink coupling **1112** from set-top box **1110**;

FIG. **20**A depicts a simplified block diagram of set-top appliance **1120** as shown in FIG. **19**C supporting coupling **1002** with hand held remote **1000** and coupling **1112** with set-top box **1110**;

FIG. **20**B depicts a simplified block diagram of set-top appliance **1120** as shown in FIG. **19**D supporting coupling **1002** with hand held remote **1000** and couplings **1112** and **1114** with set-top box **1110**;

FIG. **20**C depicts a block diagram further depicting accessibly coupled **1162** memory **1160** as shown in FIGS. **20**A and **20**B;

FIG. **21** depicts a remote control unit **1000-180** coupled **1002-180** to set-top apparatus **1100-180**, communicating via a two-stage wireline communications system containing a wireline physical transport **1200** through an augmented distributor node **1310** interfacing to a wireline communications loop including an augmented Headend **1410** further supporting a communications loop including augmented metropolitan Headend **1410**, in accordance with the invention;

FIG. **22** depicts a remote control unit **1000-180** coupled **1002-180** to set-top apparatus **1100-180**, communicating via a two-stage wireline communications system containing a wireline physical transport **1200** through a distributor node **126** interfacing to a wireline communications loop including an augmented Headend **1414** further supporting a communications loop including augmented metropolitan Headend **1410**, in accordance with the invention;

FIG. **23** depicts a detail block diagram of an augmented distributor node **1310**, coupled to wireline physical transport **1200** and coupled to the wireline communications loop of FIG. **21**;

FIG. **24** depicts an alternative detail block diagram of an augmented distributor node **1310**, coupled to wireline physical transport **1200** and coupled to the wireline communications loop of FIG. **21**;

FIG. **25** depicts a generic block diagram of a prior art Headend **104** as shown in FIG. **3**;

FIG. **26** depicts an augmented Headend **1410** of FIG. **21** or an augmented Headend **1414** of FIG. **22** or an augmented metropolitan Headend **1410** of FIGS. **21** or **22**, in accordance with the invention;

FIG. **27** depicts an alternative augmented Headend **1410** of FIG. **21** or an alternative augmented Headend **1414** of FIG. **22** or an alternative augmented metropolitan Headend **1410** of FIGS. **21** or **22**, in accordance with the invention;

FIG. **28**A depicts a block diagram of a speech engine **1330** as shown in FIG. **23**;

FIG. **28**B depicts a block diagram of a speech engine **1330** as shown in FIG. **24**;

FIG. **29** depicts a more detailed block diagram of a speech engine **1330** as shown in FIG. **28**A;

FIG. **30** depicts an alternative detailed block diagram of a speech engine **1330** as shown in FIG. **28**A;

FIG. **31** depicts a second alternative detailed block diagram of a speech engine **1330** as shown in FIG. **28**A;

FIG. **32**A depicts a block diagram of modulator engine **1350** of FIGS. **23** and **24**, and modulator engine **1450** of FIGS. **26** and **27**;

FIG. **32**B depicts a block diagram of a local oscillator as is known in the art for use as a local oscillator depicted in FIGS. **33** and **34** as *LO1* **1760**, *LO1* **1860**, *LO2* **1770** or *LO2* **1870**;

FIG. **33** depicts a detail block diagram of frequency conversion circuitry **1710** of FIG. **32**A;

FIG. **34** depicts an alternative detail block diagram of frequency conversion circuitry **1710** of FIG. **32**A; and

FIG. **35** depicts a detailed diagram of speech engine **1330** as shown in FIG. **23** or speech engine **1430** as shown in FIG. **27** containing two plex communications grids with dual, redundant gateways.

DETAILED DESCRIPTION OF THE INVENTION

FIG. **3** depicts a remote control unit **1000** coupled **1002** to set-top apparatus **1100**. Set-top apparatus **1100** communicates via a two-stage wireline communications system containing a wireline physical transport **1200** to a distributor node **1300**. Set-top apparatus **1100** communicates through distributor node **1300** across a high-speed physical transport **1400** to a tightly coupled server farm **3000**, possessing various delivery points **1510** and entry points **1512**–**1518**. Tightly coupled server farm **3000** contains one or more gateways **3100**, and one or more tightly coupled server arrays **3200**.

As used herein, a server farm refers to a collection of at least two server components communicatively coupled to one another. The server components may or may not all be directly communicatively coupled to each other. A server component refers to at least a gateway, server array, server computer, database engine, or disk farm.

As used herein, gateway refers to at least one of the following: A gateway may perform protocol conversion between different types of networks or applications; gateways may support complete conversion of one protocol to another, or support one protocol from within another protocol; a gateway may perform conversion between two messaging or mail protocols; a gateway may act as a go-between for two networks with similar protocols, providing an entry/exit point for one network in the other network. Such gateways include proxy servers; a gateway may switch speech and data between a terrestrial network and an orbiting satellite communications system; and a gateway may perform network layer switching tasks between at least two networks, coordinating the delivery of information, usually in the form of messages, packets or data streams to various destinations in the networks.

As used herein, a server array refers to a multi-dimensional lattice or array of server computers, each with an associated multi-dimensional array address, and a communication grid supporting communication between server computers based upon the multi-dimensional array addresses of the source and destination server computers. A tightly coupled server array is a server array possessing a very low latency communication grid.

The invention may include a remote control unit **1000** fitted with a microphone. Remote control unit **1000** may be fitted with such features as a special noise canceling microphone and/or a push-to-talk button.

The microphone in the remote relays the subscriber's speech commands to the central speech recognition engine. The push-to-talk button may begin the process of speech recognition by informing the system that the subscriber is about to speak and also to provide immediate address information. Address information identifies the user site at which the speaking occurs.

The invention may also include an array of microphones that are operated in conjunction with a remote control **1000** that is coupled to the set top box **1100**. The microphone array may further provide an echo-canceling capability in receiving speech signals within the area of usage.

Note that there may be more than one remote control unit **1000** with one or more microphones, each remote control unit under the control of a distinct, identified user. Such situations are particularly useful in game playing settings, where different users may wish to make commands to the game simulation like "arm the photon torpedoes", etc.

A given residence may include more than one set-top box **1100**, each of which has a distinct address in the network delivering video content and/or cable television. Each constitutes a distinct user site and may be parameterized differently. By way of example, a first set-top box in a recreation area for children may allow identified users, who are children, to select programming on only certain channels. Whereas a second set-top box in a private area of adults, such as a parental bedroom, may be parameterized so that child identifier users have no privileges.

Speech commands from the subscriber may be preprocessed. The analog signals picked up from the microphone are converted to digital signals where they undergo additional processing before being transmitted to the speech recognition and identification engine located in the cable Headend or other centralized location. Such speech preprocessing may include encryption, compression, or conversion to an alternative form of speech representation.

The preprocessing function may also take place in the remote control **1000** itself before being transmitted to the set-top box **1100**, or set-top appliance **1100**. Coupling **1002** may use a wireline or wireless physical transport. Coupling **1002** may use a wireless transport, including, but not limited to, at least one of the infra-red, microwave, or radio fre-

quency spectrum, as well as ultrasonic signaling. Coupling **1002** may support bi-directional communication between remote control **1000** and set-top box or appliance **1100**. The communication may be predominantly or strictly from remote control **1000** to set-top box or appliance **1100**.

The speech signal from the remote **1000** may be a digitally modulated RF signal whose properties may comply with Part 15 of the FCC rules. Alternatively, the set-top box **1100**, or set-top appliance **1100**, receives the speech signal from the remote **1000** and performs the preprocessing function mentioned above.

The set-top box **1100**, or set-top appliance **1100** may also be used to transmit speech and subscriber address data to the centralized location or Headend for speech recognition and identification. The Radio Frequency (RF) signal from the remote **1000** is received by the set-top appliance **1100** and then remodulated for upstream transmission **1200** on the 5 to 40 MHz cable return path. If a commercial set-top box **1100** is used to transmit the upstream speech data, then the upstream channel allocation and transmission protocol are controlled by the bi-directional communication system which is resident in the set-top box.

In certain embodiments, a commercial set-top box **1100** may not be used to transmit the digitized speech data upstream. The set-top appliance **1100** is then responsible for receiving the upstream channel allocation and synchronization information. The data receiver in the set-top appliance **1100** can be tuned to any one of several downstream data channels to receive channel and synchronization information.

The set-top appliance **1100** is also capable of receiving and decoding data in the downstream path. This function is required to organize and synchronize the transmission of upstream data which may include user feedback. Downstream data can contain upstream channel allocation information and speech verification overlay information coded as text.

The set-top box **1100** may be used for both upstream and downstream communication for the described speech command function. The function of the set-top appliance **1100** may be to receive the RF signal from the remote control and then digitize and compress the speech signal and prepare it for upstream transmission.

New RF protocol standards, such as Blue Tooth, allow the remote control's RF signal to transmit the speech signal directly to the set-top box where again, the preprocessing can either be done in the remote control **1000** or executed from firmware within the set-top box **1100**. Note that infrared signaling may be employed to transmit the speech signal to the set-top box.

Set-top boxes **1100** that employ the DOCSIS type cable modems such as Open Cable set-top boxes or the so called Heavy Set-top boxes from such companies as Scientific Atlanta and General Instruments are capable of sending and receiving speech data using efficient data transmission protocols. The DOCSIS protocol also incorporates error detection and correction capabilities as well as other transmission enhancements such as pre-equalization for more efficient and error free transmission.

FIG. **4** depicts a coupled server array **3200** of FIG. **3**.

FIG. **5** depicts a gateway **3100** of FIG. **3**.

The system uses the subscriber's address information as a mechanism by which the centrally located AgileTV™ Speech Processor can fetch a particular subscriber's parameter file. The parameter file contains speech training parameter data, speech identification parameters and user profiles for each user at that address. This file can also contain parental control information in a residential setting and other specifics for that particular address, such as language preferences or movie preferences or even internet preferences.

The Addressed Subscriber Parameter File (ASPF) gives the system an extremely high probability of speech recognition and user identification. Such addressing supports secure transactions such as banking because the speech recognition and identification system has to only identify a few parameter files for any one physical address, resulting in a very high probability of recognizing a specific speaker's speech.

Financial transaction security, e.g. speech banking and e-commerce, can be realized with this system and, with the addition of speech encryption processing in the Speech Preprocessor, even higher levels of security can be attained. This directly supports a method of contracting based upon an offer presented to the user by the system, and a recognizable acceptance of the offer by the identified user. The method of contracting may further include storing the offer and acceptance as a recording mechanism for the contract.

The speech signal transmitted from a subscriber's set-top box, or set-top appliance, **1100** is received **1510** by the five to 40 MHz data receiving equipment within gateway **3100** in the central location. As used herein, a central location may include a node, Headend, or metropolitan Headend for a residential broadband network.

When the digitized speech signal comes from a commercial set-top box, such as a General Instruments or a Scientific Atlanta set-top, then the return path receiving equipment in the central location may be specific to that type of box. Therefore, the data coming from this equipment, which contain other upstream traffic, may be parsed in such a way that only the speech commands and address information from the subscriber are input to the speech recognition engine in the central location.

When the digitized speech signal sent upstream comes from another set-top appliance, the upstream data receiver in the central location may be a separate unit that receives only speech command signals from set-top appliances at the user sites. Using the set-top appliance as the upstream transmitter allows the use of custom upstream protocols such as FM, AM, PSK, or spread spectrum digital transmission. Digital transmission techniques such as QPSK or QAM can also be employed.

Upon receiving the digitized and preprocessed speech signal from the subscriber's set-top box or set-top appliance, the received upstream signal may be in the form of a data stream containing speech and address information. The data stream containing speech and address information may be Ethernet compatible. Because the AgileTV™ Voice Processing Unit (AVPU) is a high speed speech processing unit capable of processing the data from several nodes, the digital speech signals from each of these nodes may be combined in the AVPU input multiplexer. The combining of digital speech signals may result in a smaller number of high speed digital streams.

Upstream signals such as **1510** are received at the gateway **3100**. Speech and data signals may be received from commercial return path data receivers. Speech and data signals are also received and decoded by custom return path receivers using at least one of the following protocol options: FM or AM modulation/demodulation, FDMA, TDMA, FSK, PSK, or QPSK digital modulation/demodulation, Spread Spectrum modulation/demodulation, Telephony, cellular return, or Wireless.

The AVPU Engine may not be an application service, in and of itself. The system may provide new end user appli-

cations. The AVPU Engine may provide speech recognition and control services for existing applications, such as Interactive Program Guides, Video On Demand (VOD) services or access to the Internet or World Wide Web.

Near network system initialization time, applications such as Video On Demand or Interactive Program Guides, that request service access with the speech recognition services may be required first to register with the AVPU system. At least one standard program interface may then be used to enable each application to specify its complete menu hierarchy. The invention may use exactly one standard program interface to enable applications to specify their complete menu hierarchy.

A list structure specifies each application's menu hierarchy. This list structure may be a tree. It contains labels for each menu, along with the text of each button on each menu screen, information the system needs to provide speech navigation services independently through the menu hierarchy on behalf of the application. This menu hierarchy represents the static portion of the application's data.

In addition to the static menu structure, it may also be the responsibility of the application to inform the system of dynamic content, for example the names of movies in a Video On Demand system or of program names and times in an interactive program guide. Each time a user enters a menu context in which dynamic content appears, the application may inform the speech system of this context by passing a handle associated with the list of names that comprise the dynamic content. The speech system may combine the static menu content with the augmented dynamic content (see Similarity Searching below), to form a complete grammar. Also, application-independent keywords, such as HELP, may be combined with the static menu content and augmented dynamic content to form the complete grammar. This construct may then be passed to the speech system to maximize recognition accuracy.

Given that dynamic content, by definition varies, applications may be required to inform the system whenever the content changes. In an interactive TV guide application, for example, the application registers a new set of dynamic content every one-half hour. For a VOD system, this registration may be performed whenever the database of offered movies changes.

Once registration has been completed, and the system is being used, recognition of a spoken request may cause a signal to be sent back to the application. This signal informs the application to perform the requested action, and/or to update the contents of the screen as a result of the user's request. In this manner, the application can use the system's speech recognition services with minimal modifications to the application's code, while retaining the same graphical look and feel to which users have become accustomed.

In the subscriber's premises, a speech-enabled remote control may be employed, e.g. containing a microphone, as well as traditional universal remote control functionality. The remote control may further contain a Push-To-Talk (PTT) button. Conventional remote control functions are transmitted via infra-red (IR). The speech output may be wirelessly transmitted to a set top pod, module, or appliance located at the set-top box. The wireless transmission from the remote control may employ an electromagnetic physical transport layer or ultrasonic physical transport layer. The electromagnetic physical transport layer may use a radio frequency carrier or an infra-red carrier.

When the PTT button is pushed by the user, the remote control sends a PTT active command to the set top appliance. The appliance may then inform the set-top box to place an icon on the screen, or otherwise indicate to the user that the system is listening to them. Next, as the user speaks into the microphone, the speech is digitized, compressed, and transmitted to the set top appliance.

The set top appliance may perform any or all of the following: encrypt the speech sample to provide security; add subscriber address information; add a message length code; add error control coding to the assembled information. The error control coding may include a Cyclical Redundancy Code (CRC), enabling data transmission errors to be detected.

In homes with set-top boxes having the necessary features, the set top appliance may transmit speech information to the set-top box, which then transmits it to the central location as a series of packets.

Alternatively, the set top appliance may directly transmit the speech stream to the central location itself. This may continue until the set top appliance receives a PTT Release from the remote, indicating end of speech. This information is also transmitted to the central location, signaling end of spoken request.

Each individual consumer's interface, i.e. set-top box or set-top appliance, may have a unique address. The address of the individual consumer's interface may be determined during the manufacturing process. As speech packets are transmitted upstream, this address information may be appended to the speech packets. The address information enables rapid determination of the user site from which the speech sample is received. The address information may precede the speech packets. The address information improves the efficiency of at least some of the central location processing stages.

Assignment of an input buffer address to the speech sample may be based upon the user site address. This input buffer collects incoming speech packets until the final packet of a spoken request has been received. Recognition of the final speech packet receipt may be encoded in that speech packet. Final speech packet receipt recognition may include a delay in speech of greater than a predetermined amount of time. The speech delay amount may be one or more of at least the following: a globally defined constant, a user site defined constant, an identified user defined constant, or an application-defined constant for an application serviced by the speech recognition system. The final speech packet may be generated based upon the release of the PTT button in the remote control unit. The final speech packet may be generated at the remote control unit, set top pod, set top appliance, set top box, node, Headend or metropolitan Headend, or by the serviced application. The final speech packet may be generated by the serviced application based upon a time schedule, e.g. a fixed period in which participants may respond, a privilege scheme, and/or other application related scheme.

On-the-fly Cyclical Redundancy Code (CRC) error checking generation may be used. Each time a packet is read in, CRC in the processor registers may be computed as each byte is read, then the partial CRC is stored with the stored packet. The CRC may be stored at the end of the packet, at the beginning of the packet or in specifically designated locations within the packet. When the next packet arrives, the partial CRC is read from where it was stored, and the new packet data may be appended to the end of the previous packet, overwriting the temporary CRC. This continues until a complete speech sample has been received. This approach to CRC calculation cuts memory accesses in half compared to first storing the string, then making a second pass to generate the CRC.

Once a complete spoken request has been received, the speech input processor may use the sample's source address identifying a user site to target the speech data to a specific speech processing processor. This direct correspondence between the source address and a specific speech processor allows speech processors to cache user-specific parameters for the user sites they serve. The mapping of source addresses to speech processor reduces the bandwidth necessary to move user site-specific data to each speech processor. In other embodiments of the invention, a translation table may be used to allow speech to processor assignments to be changed dynamically in the event of a hardware or other failure, e.g. capacity overload, while retaining the efficiency advantages of directly mapping the speech channel to the processor.

To process a speech sample by the speech processor into speech content, the speech processor must first contain in its local memory a copy of the grammar definition associated with the user site. A grammar is a structure often containing the words which are most likely to be spoken, the order in which these words may appear, and the meaning of various sequences of words. This structure may be a hardware configuration, software program, data structure, or a combination of two or more of these.

Before transferring the new speech sample to a speech processor, the grammar associated with the speech sample is transferred to the target speech processor. The grammar transfer may use a simple Least Recently Used (LRU) queue. If the speech processor contains enough empty space in its memory, then the indicated grammar is transferred directly to the empty memory from mass storage. Mass storage may include, but is not limited to, a disk, disk farm, or Redundant Array of Independent Disks (RAID), a high bandwidth disk farm. If not, then the least-recently-used grammar entry may be discarded, and the new grammar information loaded into the vacated memory.

The next step in processing the speech sample ensures that the current parameters associated with this user site are already cached in the specific speech processor's RAM. If these parameters are not present, then the least-recently-used parameter cache entry may be evicted from the cache.

To do this, the oldest cache entry on this speech processor is first examined to see if it has been modified. If it has, the cache entry is stored to mass storage, and the cache slot is then declared vacant. Next, the user site speech parameters associated with the new speech sample are loaded into the vacated cache block. During the relatively long access times needed to load a new set of user site parameters from mass storage (and optionally to write the old parameters to the disk. etc.), the current speech sample may be held in the input buffer in a waiting state.

After the new household speech parameters are loaded into the targeted speech processor, the speech sample is moved into the work queue for the speech processor. This minimizes speech processor delay for processing other speech requests during disk accesses. Instead, the speech processor may process other speech samples associated with user sites whose parameters are already in the cache.

Once the speech parameters associated with a speech sample are available in the speech processor, the speech sample may be assigned to the speech processor by placing a description of the speech sample into the target speech processor's work queue. As speech samples are processed, they may be removed from the front of the work queue by the speech processor.

Eventually, the speech processor reaches the current input sample. The speech sample is transferred into the speech processor's local memory, and the status of this speech sample changed to Next. This transfer may occur in parallel with the processing of the prior speech sample, ensuring that speech processor utilization is maximized.

Once this transfer is complete, and the speech processor completes processing of the prior sample, the status of this speech sample changes to Current, and the speech recognition engine begins processing this sample.

Successful processing of a speech sample requires that both the proper grammar and the proper user site parameter information be simultaneously accessible by the speech processor.

The possibility exists in which a recently-loaded grammar or speech parameter file is erased prior to its use, in the process of loading the grammar or speech parameters for the current speech sample. To eliminate this condition, the total number of speech samples sitting in the waiting and working queues of a speech processor preferably should not exceed the number of cache entries in the speech processor.

The first step in recognizing the current speech sample may determine which individual person pronounced the current spoken request. The Speaker Identification software module running on the targeted speech processor compares the spoken characteristics of this speech sample with the characteristics of the speakers who have been previously identified in this user site.

In many spoken requests, the incoming speech sample matches the characteristics of a previously identified speaker. When this occurs, the speech sample is passed on to the next phase, speech recognition.

If the speech sample is not identified with an existing speaker, then a new user routine is invoked, enabling a new user to be associated with this household. This routine records the new individual's speech parameters into this user site's speech parameters, so that the new speaker may be identified during subsequent spoken requests. Optionally, the system may offer the user site manager the opportunity to erase or modify the new speaker parameters.

The inputs to the speech recognition software module may include a speech sample, an individual user's speech parameters, and the grammar to be recognized. The speech engine determines the most likely spoken request based on statistical analysis, and may return a text string corresponding to the spoken request. This matching process is probabilistic: along with the returned text string, the speech engine may also return a confidence measure, or percentage of match likelihood. The returned text string and confidence measure enables different applications to respond differently based on the calculated confidence in the recognition result.

For recognition results having a low cost to the user, such as a request to display listings for a particular movie, lower confidence criteria may be applied. For recognition results with a high cost to the user, such as a request to purchase a movie, higher confidence thresholds may be required. Furthermore, purchase verification may be requested.

When recognition accuracy is particularly low, the speech recognition engine may determine partial matches to more than one possible phrase, returning the text of several possible matches. This process enables an application or user to select from several alternative recognition results.

In cases where a transaction results in a charge to the user, the system may record the speech sample representing the purchase request onto mass storage. This speech sample recording is performed to provide verification that a purchase request was made, should a user dispute their intent to purchase. This further supports speech-based methods of contracting. Note that the recording may be compressed in

either a standard or non-standard manner, which may further include encryption technology.

Due to the statistical nature of speech recognition, spoken requests are occasionally mis-recognized. Recognition errors have been a primary hindrance to further exploitation of speech recognition systems. This system provides a mechanism by which the user receives rapid visual feedback regarding the recognition process. Soon after the speech recognition engine has returned a result, visual text corresponding to the recognized spoken request is displayed on the display, e.g. television, screen.

This rapid visual feedback may be accomplished by transmitting the recognized text string back to the set-top box. Software executing within the set-top box displays the text information in a special window on top or overlaying of the existing application display. In cases where limitations in the set-top box prevent the window from appearing on top of existing content, software executing within the set-top box may select an alternative feedback display, for example, a different channel containing a background field, and displays the overlay information on top of this background.

In cases where the recognition accuracy is particularly poor, and the speech engine returns several possible recognition results, this overlay display capability may be used to help refine the user's query. By displaying the text of the possible recognition results, the user can easily select from the returned list.

If none of the recognized text strings match the user's spoken request, the user may elect to start again. However, more commonly, one of the displayed strings match, or be likely to readily lead to the user's speech request with appropriate modification. When this is the case, the user can readily select from the displayed alternatives, enabling improved productivity in low-recognition situations.

One key feature which can be used with applications such as Pay-Per-View and Video-On-Demand, is similarity searching. When the system is scanning for dynamic content, i.e. movie titles and actor names, a similarity search feature may be available. This enables searches for names which are only partially matched, or which resemble the recognized phrase, without requiring precise specification of the exact title or name. The technique used to deliver similarity searching is to augment the dynamic content list specified by the application. Words similar in meaning, content, or spelling to words in the content list may be automatically added to the dynamic content list whenever the application informs the system of a content update. This enables the recognition grammar to be extended to support a broader array of matches without requiring bandwidth-intensive text searches each time a user's speech request has been received. Links between similar and target words may be maintained in the internal representation of the grammar, thereby providing direct access to target names. This further enhances system performance.

Note that as used herein, similarity databases may be constructed from at least one or more of the following database architectures: relational, SQL, knowledge-based, inferential knowledge bases including Horn clause and extended Horn clause fact bases, fuzzy versions of the above, as well as neural networks and combinations of these techniques. Also, manual additions may be made at the AgileTV™ central office.

The entire system may provide fault tolerance for robust performance. For individual components, such as speech processors and associated memory, the system may automatically recognize most component failures, and disable faulty processors.

In addition, many system installations may require multiple AgileTV™ Voice Processing Unit (AVPU) boxes **3000** to meet the performance needs of the subscriber base. To ensure maximum operating time, even the total failure of an AVPU engine may not be catastrophic. The remaining AVPU engines continue to service the incoming speech traffic at a reduced performance level. This failover is carried out by a negotiation protocol between the AVPU boxes.

In addition to these internal diagnostics functions, the system may also communicate continuously with a network operations center, enabling rapid identification of system errors, as well as rapid service dispatch to correct any problems.

A speech command preprocessor may perform the function of speech filtering, digitizing, data compression, encoding pauses in speech, and address insertion. Preprocessing speech commands at user sites results in a lower upstream data rate. This processing can be located in the remote **1000**, the Set-top appliance **1100**, or the set top box.

A Speech Packet Processor may be centrally located in or near a wireline node specifically to capture and prepare the upstream speech packets that are to be fed to the Speech Recognition Engine.

A speech processor system may be centrally located in or near a wireline node, which may include a Cable Television (CATV) central location. The speech recognition system may be centrally located in or near a server farm. The speech recognition system may be centrally located in or near a web-site hosting location. The speech recognition system may be centrally located in or near a gateway.

The speech recognition engine processes speech packets to create speech content and formulate the response to the speech content for each of the user sites.

The system also performs address decoding for routing and latency reduction. It also performs the function of decoding the pauses between words or sounds originally encoded on the preprocessor at the user site.

A method is provided for a very high level of security based on the ability of being able to precisely recognize a parent's speech or a child's speech based on a limited subset of choices. The ability to recognize an individual speaker with a very high probability of identifying the specific speaker is based on knowing the address of the user site.

A large capacity, centrally located, natural speech recognition engine can readily identify a specific speaker from among a small set of choices, allowing a parent to voice print and therefore control the type, or amount or timing of content a child may request. A child may be able to learn PIN numbers but cannot readily fool the voice print. Speaker recognition is particularly appealing in this parental control application, and can be applied with very high accuracy, due to the limited number of speakers per user site. Parents may control the specific viewing times, total time, or content that a child can watch television through this system and its use.

The operational discussion just presented regarding FIGS. **3–5** is applicable to the relevant operational system embodiments based upon any and all of the FIGS. **19**A to **24**, as well as, FIGS. **26** to **31**.

A parallel processor system supporting this sort of system is disclosed in U.S. Patent Application Ser. No. 60/210,440, entitled "System and Method of a Multi-dimensional Plex Communication Network", which describes plex processor networks and which is incorporated herein by reference.

A plex communications network has M orthogonal directions that support communications between an M dimensional lattice that may include N^M plex-nodes, where M is at least two and N is at least four. Each plex-node pencil in

a first orthogonal direction contains at least four plex-nodes and each plex-node pencil in a second orthogonal direction contains at least two plex-nodes. Each of the plex-nodes contains a multiplicity of ports.

A plex-node pencil refers to a 1-dimensional collection of plex-nodes differing from each other in only one dimensional component, i.e. the orthogonal direction of the pencil. By way of example, a nodal pencil in the first orthogonal direction of a two-dimensional array contains the plex-nodes differing in only the first dimensional component. A nodal pencil in the second orthogonal direction of a two-dimensional array contains the plex-nodes differing in only the second dimensional component.

The plex communications network is comprised of a communication grid interconnecting the plex-nodes. The communications grid may include $N^*(M-1)$ communication pencils, for each of the M directions. Each of the communication pencils in each orthogonal direction is coupled with a corresponding plex-node pencil containing a multiplicity of plex-nodes coupling every pairing of plex-nodes of the corresponding plex-node pencil directly.

Communication between two plex-nodes of a nodal pencil coupled with the corresponding communication pencil comprises traversal of the physical transport layer(s) of the communication pencil.

Such plex communications networks support direct communication between any two plex-nodes belonging to the same communication pencil. This supports communication between any two plex-nodes in at most M hops between plex-nodes.

FIG. **6** depicts a tightly coupled server farm **3000** of FIG. **3** implemented in a two-dimensional plex communication network with N=4 plex-nodes in each of two orthogonal directions of the plex-node array.

Consider the plex-node pencil in the first orthogonal direction containing Plex-node 0,0, Plex-node 1,0, Plex-node 2,0 and Plex-node 3,0. The communication pencil **400** in the first orthogonal direction couples to the plex-nodes of this plex-node pencil. Plex-node 0,0 is coupled **402** to communication pencil **4100**. Plex-node 1,0 is coupled **404** to communication pencil **400**. Plex-node 2,0 is coupled **406** to communication pencil **400**. Plex-node 3,0 is coupled **408** to communication pencil **400**.

Consider the plex-node pencil in the first orthogonal direction containing Plex-node 0,1, Plex-node 1,1, Plex-node 2,1 and Plex-node 3,1. The communication pencil **410** in the first orthogonal direction couples to the plex-nodes of this plex-node pencil. Plex-node 0,1 is coupled **412** to communication pencil **410**. Plex-node 1,1 is coupled **414** to communication pencil **410**. Plex-node 2,1 is coupled **416** to communication pencil **410**. Plex-node 3,1 is coupled **418** to communication pencil **410**.

Consider the plex-node pencil in the first orthogonal direction containing Plex-node 0,2, Plex-node 1,2, Plex-node 2,2 and Plex-node 3,2. The communication pencil **420** in the first orthogonal direction couples to the plex-nodes of this plex-node pencil. Plex-node 0,2 is coupled **422** to communication pencil **420**. Plex-node 1,2 is coupled **424** to communication pencil **420**. Plex-node 2,2 is coupled **426** to communication pencil **420**. Plex-node 3,2 is coupled **428** to communication pencil **420**.

Consider the plex-node pencil in the first orthogonal direction containing Plex-node 0,3, Plex-node 1,3, Plex-node 2,3 and Plex-node 3,3. The communication pencil **430** in the first orthogonal direction couples to the plex-nodes of this plex-node pencil. Plex-node 0,3 is coupled **432** to communication pencil **430**. Plex-node 1,3 is coupled **434** to communication pencil **430**. Plex-node 2,3 is coupled **436** to communication pencil **430**. Plex-node 3,3 is coupled **438** to communication pencil **430**.

Three of these ports on each plex-node are devoted to providing a direct interconnect to the other plex-nodes of its row through a collection of communication paths forming the communication pencil in the second orthogonal direction. These plex-nodes belong to the same row as the plex-nodes of the plex-node pencil in the second orthogonal direction.

A plex-node may have at least one additional port. At least one of the additional ports may be connected to an external network. Further, at least one of the additional ports may be connected to an external mass storage system. In other embodiments of the invention, at least one of the additional ports may be connected to an external database system.

A plex-node may contain at least one instruction processor. As used herein, an instruction processor includes but is not limited to instruction set processors, inference engines and analog processors. An instruction set processor refers to instruction processors changing state directly based upon an instruction, and which change an internal state by executing the instruction. The instruction may include, but is not limited to, direct or native instructions and interpreted instructions. An inference engine changes state when presented an instruction, which may include an assertion, an assumption, or an inference rule. Inference engines include, but are not limited to, Horn clause engines such as Prolog requires, constraint based systems and neural network engines. As referred to herein, analog processors include, but are not limited to, optical signal processors, CCD's, and resonant cavity devices responding to data and/or controls asserted in the analog domain.

Communication includes, but is not limited to, communication using a digital communications protocol. Communication also includes a messaging protocol using the digital communications protocol. Communications also includes a messaging protocol supporting TCP-IP, supporting the Internet, and/or supporting the World Wide Web.

Communications may also include at least one video stream protocol using a digital communications protocol. Communications may include at least one multi-media stream protocol using the video stream protocols including motion JPEG and/or at least one form of MPEG.

The invention may support Wavelength Division Multiplex (WDM) through the physical transport of the communication paths of the communication pencils.

Each plex-node may include a communication processor. Each plex-node may further include M communications processors. The ports of each communications processor may be connected to each one of the pencils supporting pencil communications processing.

FIG. **7** depicts a gateway **3100** of FIG. **3** implemented in a two-dimensional plex communication network with N=4 plex-nodes in each of two orthogonal directions of the plex-node array.

In FIGS. **6** and **7**, upstream signals **1510** may be received at the Plex-node 0,0, performing functions similar in nature to processor node **3110** of FIGS. **3**, **4** and **5**. Plex-node 0,1 may couple **3142** to RAID **3130** subsystem, providing high-speed access to a high capacity disk farm. Various signals **1512** to **1518** may be generated by various plex-nodes. Note that because of the uniformity of the communication structure, there is great flexibility in choosing which plex-node generates which communication signals.

FIG. **7** also shows coupling **3002** to a speech recognition server array **3200**.

FIG. **8** depicts a coupled server array **3200** of FIG. **3** implemented in a two-dimensional plex communication network with N=4 plex-nodes in each of two orthogonal directions of the plex-node array.

Each plex-node of FIG. **8** may include a plex communications network supporting a two or more dimensional array of internal plex-nodes, each including at least one instruction processor.

FIG. **9** depicts a simplified block diagram using an array of processors as shown as **3200** in FIG. **3**.

Consider an array of plex-nodes, where each plex-node uses multiple-processor IC's labeled CPU. Each CPU may possess three or more high speed serial protocol channels C1-3, a high speed interconnect bus B1, a low speed interconnect bus PCI and a local memory access bus R. Each CPU may include at least two processors. Each processor may be an instruction processor. Each CPU may be an integrated circuit. The integrated circuit may be a BCM12500 manufactured by Broadcom Corporation of Irvine, Calif.

CPU1 and CPU2 are locally coupled via at least one of the interconnect buses. Each CPU possesses locally accessible memory via an R bus. Each CPU may further access locally accessible memory via its own R bus. The R bus accessed memory may include DDR SDRAM (Double Data Rate Synchronous DRAM).

Three serial protocol channels for the first CPU of each CPU pair are dedicated to communication within each row of CPU pairs. Three serial protocol channels for the second CPU of each CPU pair are dedicated to communication within each column of CPU pairs.

Assume each CPU contains at least two processors. Assume each processor is operating at K billion integer operations/second and L billion floating point operations/second. Assume the high speed serial channels support a gigabit/second bidirectional transfer rates. The DDR DRAMs support M gigabyte/second access rates. This system provides 64*K billion integer operations, 64*L billion floating point operations, 128 gigabits/second interconnect communication bandwidth throughout the array through the high speed serial channels, as well as 64*M gigabytes/second local memory access. By decoupling the tasks running on each processor, making them independent and concurrent, this system architecture achieves the dream of Multiple-Instruction Multiple Datapath computing (MIMD), providing essentially linear speed up with the increase in processors for up to 64 processors.

Such architectures can support each processor generating a frame in a local frame buffer and rendering it into a motion video stream, which is merged into a multiplexed video content stream suitable for a variety of multi-media applications. This apparatus meets or exceeds the requirements of a number of teraflop-class computing applications, which today have no cost effective solution. Note that the processors may each be dual instruction processors housed in a single package.

Each of these arrays will process over 1K speakers concurrently and fit into a single rack mount enclosure. Standard 72 inch, tall racks could then hold 4 such speech engines, also known herein as an AVPU, supporting up to 140,000 subscribers. Each server contains at least dual, redundant, hot-swappable power supplies, and redundant, hot-swappable disks with automatic recovery for failed disk drives. The hot-swappable power supplies may further be half-sized, each supporting the power requirements of half the server engine.

Real-time, extensive diagnostics and continuous connectivity to a 24 hours per day, seven days per week (24x7) Network Operations Center insuring timely failure detection and service dispatch. In systems with more than one speech engine, the server automatically handles catastrophic failures gracefully. Even if an entire speech engine fails, the remaining ones pick up the workload.

FIG. **10** depicts a flowchart of a method that uses a back channel containing a multiplicity of identified speech channels for speech recognition at a wireline node in a network supporting video or cable television delivery in accordance with the invention. The back channel is from a multiplicity of user sites and is presented to a speech processing system at the wireline node in the network. The speech processing system performs the operations of the method.

Operation **2000** starts the operations of this flowchart. Arrow **2002** directs the flow of execution from operation **2000** to operation **2004**. Operation **2004** performs receiving the back channel to create a received back channel. Arrow **2006** directs execution from operation **2004** to operation **2008**. Operation **2008** terminates the operations of this flowchart.

Arrow **2010** directs the flow of execution from starting operation **2000** to operation **2012**. Operation **2012** performs partitioning the received back channel into a multiplicity of received identified speech channels. Arrow **2014** directs execution from operation **2012** to operation **2008**. Operation **2008** terminates the operations of this flowchart.

Arrow **2020** directs the flow of execution from starting operation **2000** to operation **2022**. Operation **2022** performs processing the multiplicity of the received identified speech channels to create a multiplicity of identified speech content. Arrow **2024** directs execution from operation **2022** to operation **2008**. Operation **2008** terminates the operations of this flowchart.

Arrow **2030** directs the flow of execution from starting operation **2000** to operation **2032**. Operation **2032** performs responding to the identified speech content to create an identified speech content response, for each of the multiplicity of the identified speech contents. Arrow **2034** directs execution from operation **2032** to operation **2008**. Operation **2008** terminates the operations of this flowchart.

Portions of the flowchart of FIG. **10** related to operations **2022** and **2032** may operate a speech processing system coupled to a wireline node in a network. The wireline node provides multiple received identified speech channels to the speech processing system. The received identified speech channels are based upon a received back channel at the wireline node from multiple user sites coupled to the network. The network supports video delivery to the user sites and/or cable television delivery to the user sites.

The speech processing system may include at least one computer. Operations **2022** and **2032** may be implemented as program steps of a program system operating at least part of the computers included in the speech processing system. Implementations of these operations as program steps may reside in memory accessibly coupled to at least one computer in the speech processing system.

Note that operations **2002** and **2012** may be implemented in hardware and/or implemented as program steps used in operating computers located at the wireline node. Such computers may or may not be part of the speech processing system.

At least one, and possibly all, of the identified speech channels may have an associated user site.

FIG. **11A** depicts a detail flowchart of operation **2012** of FIG. **10** further partitioning the received back channel.

Arrow **2070** directs the flow of execution from starting operation **2012** to operation **2072**. Operation **2072** performs partitioning the received back channel into a multiplicity of received identified speech channels from the associated user site. Arrow **2074** directs execution from operation **2072** to operation **2076**. Operation **2076** terminates the operations of this flowchart.

FIG. **11**B depicts a detail flowchart of operation **2022** of FIG. **10** further processing the multiplicity of the received identified speech channels.

Arrow **2090** directs the flow of execution from starting operation **2022** to operation **2092**. Operation **2092** performs processing the multiplicity of the received identified speech channels from the associated user site to create multiple identified speech contents. Arrow **2094** directs execution from operation **2092** to operation **2096**. Operation **2096** terminates the operations of this flowchart.

FIG. **11**C depicts a detail flowchart of operation **2032** of FIG. **10** further responding to the identified speech content.

Arrow **2110** directs the flow of execution from starting operation **2032** to operation **2112**. Operation **2112** performs responding to the identified speech content from the associated user site to create the identified speech content response for the associated user site. Arrow **2114** directs execution from operation **2112** to operation **2116**. Operation **2116** terminates the operations of this flowchart.

FIG. **12** depicts a detail flowchart of operation **2000** of FIG. **10** further performing the method using the back channel from multiple user sites.

Arrow **2120** directs the flow of execution from starting operation **2000** to operation **2122**. Operation **2122** performs determining the associated user site from the received identified speech channel. Arrow **2124** directs execution from operation **2122** to operation **2126**. Operation **2126** terminates the operations of this flowchart.

Arrow **2130** directs the flow of execution from starting operation **2000** to operation **2132**. Operation **2132** performs determining the associated user site from the identified speech content. Arrow **2134** directs execution from operation **2132** to operation **2126**. Operation **2126** terminates the operations of this flowchart.

Arrow **2140** directs the flow of execution from starting operation **2000** to operation **2142**. Operation **2142** performs determining the associated user site from the identified speech content and a speaker identification library. Arrow **2144** directs execution from operation **2142** to operation **2126**. Operation **2126** terminates the operations of this flowchart.

Arrow **2150** directs the flow of execution from starting operation **2000** to operation **2152**. Operation **2152** performs determining the associated user site from the identified speech content and a speech recognition library. Arrow **2154** directs execution from operation **2152** to operation **2126**. Operation **2126** terminates the operations of this flowchart.

Arrow **2160** directs the flow of execution from starting operation **2000** to operation **2162**. Operation **2162** performs determining the associated user site from an identification within the speech channel. Arrow **2164** directs execution from operation **2162** to operation **2126**. Operation **2126** terminates the operations of this flowchart.

Note that the determination of the associated user site may be provided by an identification within the speech channel. By way of example, a technician may be recognizable at many user sites, and may identify the user site as Room 432 or 10 Main Street in the process of activities at that user site.

The invention can include at least one of the operations **2120**, **2132**, **2142**, **2152** and **2162**.

FIG. **13**A depicts a detail flowchart of operation **2112** of FIG. **11**C further responding to identified speech contents.

Arrow **2170** directs the flow of execution from starting operation **2112** to operation **2172**. Operation **2172** performs processing the identified speech content response to create the identified user site response. Arrow **2174** directs execution from operation **2172** to operation **2176**. Operation **2176** performs sending the identified user site response to the identified user site. Arrow **2178** directs execution from operation **2176** to operation **2180**. Operation **2180** terminates the operations of this flowchart.

FIG. **13**B depicts a detail flowchart of operation **2112** of FIG. **11**C further responding to the identified speech content.

Arrow **2190** directs the flow of execution from starting operation **2112** to operation **2192**. Operation **2192** performs assessing the speech content response identified as to the user site to create a financial consequence identified as to the user site. Arrow **2194** directs execution from operation **2192** to operation **2196**. Operation **2196** performs billing the user site based upon the financial consequence. Arrow **2198** directs execution from operation **2196** to operation **2200**. Operation **2200** terminates the operations of this flowchart.

FIG. **14** depicts a detail flowchart of operation **2112** of FIG. **11**C further responding to the identified speech content from the associated user site.

Arrow **2210** directs the flow of execution from starting operation **2112** to operation **2212**. Operation **2212** performs assessing the speech response to create a financial consequence identified as to the user site. Arrow **2214** directs execution from operation **2212** to operation **2216**. Operation **2216** performs displaying the financial consequence to create a displayed financial consequence at the user site. Arrow **2218** directs execution from operation **2216** to operation **2220**. Operation **2220** performs confirming the displayed financial consequence from the user site to create a financial commitment. Arrow **2222** directs execution from operation **2220** to operation **2224**. Operation **2224** performs billing the user site based upon the financial commitment. Arrow **2226** directs execution from operation **2224** to operation **2228**. Operation **2228** terminates the operations of this flowchart.

FIG. **15**A depicts a detail flowchart of operation **2112** of FIG. **11**C further responding to the identified speech content from the associated user site.

Arrow **2250** directs the flow of execution from starting operation **2112** to operation **2252**. Operation **2252** performs identifying a user based upon the speech content identified with the user site to create an identified user. Arrow **2254** directs execution from operation **2252** to operation **2256**. Operation **2256** terminates the operations of this flowchart.

FIG. **15**B depicts a detail flowchart of operation **2252** of FIG. **15**A further identifying the user.

Arrow **2270** directs the flow of execution from starting operation **2252** to operation **2272**. Operation **2272** performs fetching a user profile list based upon the user site and the user profile list containing at least one user profile. Arrow **2274** directs execution from operation **2272** to operation **2276**. Operation **2276** performs identifying the user based upon the speech content and based upon the user profile list to create an identified user. Arrow **2278** directs execution from operation **2276** to operation **2280**. Operation **2280** terminates the operations of this flowchart.

FIG. **16** depicts a detail flowchart of operation **2112** of FIG. **11**C further responding to the identified speech content from the associated user site.

Arrow **2290** directs the flow of execution from starting operation **2112** to operation **2292**. Operation **2292** performs assessing the speech response based upon the identified user

to create a financial consequence for the identified user. Arrow **2294** directs execution from operation **2292** to operation **2296**. Operation **2296** performs displaying the financial consequence to create a displayed financial consequence at the user site. Arrow **2298** directs execution from operation **2296** to operation **2300**. Operation **2300** performs confirming the displayed financial consequence from the user site based upon the identified user to create a financial commitment. Arrow **2302** directs execution from operation **2300** to operation **2304**. Operation **2304** performs billing the identified user based upon the financial commitment. Arrow **2306** directs execution from operation **2304** to operation **2308**. Operation **2308** terminates the operations of this flowchart.

FIG. **17**A depicts the speech content response **2350**, including the current response menu **2352** and the cumulative user site response **2354**, in accordance with the invention.

The speech content response may include a current response menu and a cumulative user site response identified as to the user site for at least one of the user sites.

FIG. **17**B depicts a detail flowchart of operation **2112** of FIG. **11**C further responding to the identified speech content from the associated user site.

Arrow **2330** directs the flow of execution from starting operation **2112** to operation **2332**. Operation **2332** performs responding to the speech contents based upon the current response menu and based upon the cumulative user site response to create a new cumulative user site response. Arrow **2334** directs execution from operation **2332** to operation **2336**. Operation **2336** terminates the operations of this flowchart.

Arrow **2340** directs the flow of execution from starting operation **2112** to operation **2342**. Operation **2342** performs assessing the speech contents based upon the current response menu and based upon the cumulative user site response to create a new current response menu. Arrow **2344** directs execution from operation **2342** to operation **2336**. Operation **2336** terminates the operations of this flowchart.

At least one of operations **2332** or **2342** may be performed.

FIG. **18**A depicts a detail flowchart of operation **2112** of FIG. **11**C further responding to the identified speech content from the associated user site.

Arrow **2370** directs the flow of execution from starting operation **2112** to operation **2372**. Operation **2372** performs responding to the speech content, identified as to the user site, based upon a natural language, to create a speech content response of the speech content identified as to user site. Arrow **2374** directs execution from operation **2372** to operation **2376**. Operation **2376** terminates the operations of this flowchart.

FIG. **18**B depicts a detail flowchart of operation **2092** of FIG. **11**B further processing the multiplicity of the received speech channels.

Arrow **2390** directs the flow of execution from starting operation **2092** to operation **2392**. Operation **2392** performs processing the received speech channels from the user site based upon a natural language for the user site to create the speech content identified as to the user site. Arrow **2394** directs execution from operation **2392** to operation **2396**. Operation **2396** terminates the operations of this flowchart.

FIG. **19**A depicts a simplified block diagram of a hand held remote **1000**, containing microphone **1060** and keypad **1020** supporting user input organized and processed by embedded controller **1050** for communication by wireless interface **1040** coupled **1002** to set-top apparatus **1100**, as shown in FIG. **3**.

FIG. **19**B depicts a simplified block diagram of set-top apparatus **1100** as shown in FIG. **3** showing coupling **1002** and first wireline physical transport **1200** comprised of downlink coupling **1202** and uplink coupling **1204**.

Note that microphone **1060** may include more than one audio-sensor and/or a microphone array of two or more microphone units.

FIG. **19**C further depicts set-top apparatus **1100** as shown in FIG. **19**B containing a set-top appliance **1120** coupled **1002** with hand held remote **1000**. Set-top appliance **1120** is coupled **1112** with set-top box **1110** possessing downlink coupling **1202** and uplink coupling **1204**.

FIG. **19**D further depicts set-top apparatus **1100** as shown in FIG. **19**B containing a set-top appliance **1120** coupled **1002** with hand held remote **1000** and possessing downlink coupling **1202** and uplink coupling **1204**. Set-top appliance **1120** provides processed downlink coupling **1114** to set-top box **1110** and receives initial uplink coupling **1112** from set-top box **1110**.

Regarding FIG. **19**A–**19**D, the following observations may be made.

The invention supports uni-directional communication via coupling **1002**, supporting communicative transfer from the remote **1000** via coupling **1002** to set-top apparatus **1100**.

The invention supports bi-directional communication via coupling **1002**. Note that noise-canceling microphones **1060** may use bi-directional communication on coupling **1002**. Noise cancellation may be performed within set-top apparatus **1100**, by at least one of the set-top box **1110** and set-top appliance **1120**.

Wireless interface **1040** interacts with coupling **1002**. Coupling **1002** may use a wireless transport, including, but not limited to, at least one of infra-red, microwave or radio frequency spectrum regions, as well as ultrasonic signaling. Embedded controller **1050** controls **1042** wireless interface **1040**. Embedded controller **1150** communicates via **1042** with wireless interface **1040** to direct communication across coupling **1002**.

FIG. **20**A depicts a simplified block diagram of set-top appliance **1120** as shown in FIG. **19**C supporting coupling **1002** with hand held remote **1000** and coupling **1112** with set-top box **1110**.

Computer **1150** is coupled **1132** to remote interface **1130** and also coupled **1142** to set-top box interface **1140**. Computer **1150** executes a program system including program steps residing in accessibly coupled **1162** memory **1160**.

The program system executed by computer **1150** includes program steps maintaining hand held remote **1000** communication through remote interface **1130** and maintaining set-top box **1120** communication through set-top box interface **1140**.

Remote interface **1130** couples **1002** with hand held remote **1000** (not shown in FIG. **20**A).

Remote interface **1130** incorporates the necessary device or devices to provide communications using the embodiment's physical transport layer.

Set-top box interface **1140** couples **1112** with set-top box **1120**, as shown in FIG. **19**C. Coupling **1112** may use a wireline or wireless physical transport. Coupling **1112** may use a wireless transport, including, but not limited to, at least one of the infra-red, microwave or radio frequency spectrum, as well as ultrasonic signaling. Set-top box interface

1140 incorporates the necessary device or devices to provide communications using the embodiment's physical transport layer.

FIG. 20B depicts a simplified block diagram of set-top appliance 1120 as shown in FIG. 19D supporting coupling 1002 with hand held remote 1000 and couplings 1112 and 1114 with set-top box 1110.

Set-top box interface 1140 provides processed downlink coupling 1114 to set-top box 1110 and receives initial uplink coupling 1112 from set-top box 1110 as shown in FIG. 19D. Set-top box interface 1140 incorporates the necessary device or devices to provide these communications.

FIG. 20C depicts a block diagram further depicting accessibly coupled 1162 memory 1160 as shown in FIGS. 20A and 20B.

Accessibly coupled 1162 memory 1160 may contain RAM memory 1180 coupled 1182 to computer 1150. Accessibly coupled memory 1160 may include more than one RAM memory 1180. Distinct RAM memories 1180 may be distinctly coupled to computer 1150. One or more RAM memories 1180 may act as a cache memory for computer 1150.

Accessibly coupled 1162 memory 1160 may contain non-volatile memory 1190 coupled 1192 to computer 1150. Accessibly coupled memory 1160 may include more than one non-volatile memory 1190. Distinct non-volatile memories 1190 may be distinctly coupled to computer 1150. A non-volatile memory may be organized as file management system.

Note that the coupling 1182 of RAM memory 1180 may be distinct from coupling 1192 of non-volatile memory 1190 with computer 1150. Either RAM memory and/or non-volatile memory components may be packaged with computer 1150.

FIG. 21 depicts a remote control unit 1000-180 coupled 1002-180 to set-top apparatus 1100-180. Set-top appliance 1100-180 communicates via a one- or two-stage wireline communications system containing a wireline physical transport 1200 to an augmented distributor node 1310. The communications traverse augmented distributor node 1310 interfaced to a wireline communications loop including an augmented Headend 1410. The network may further support a communications loop including augmented metropolitan Headend 1410.

As used herein, the adjective augmented is used to refer to a node incorporating at least one embodiment of the invention.

Augmented node 1310 may control and support optimized upstream communication as disclosed in the co-pending application Ser. No. 09/679,115, entitled "Increased Bandwidth in Aloha-based Frequency Hopping Transmission Systems" by Calderone and Foster, both inventors of this application and commonly assigned to AgileTV, and incorporated herein by reference.

Augmented node 1310 may provide cached content such as popular videos for a VOD service.

Speech processing may be performed in many situations at augmented node 1310.

Augmented Headend 1410 and augmented metropolitan Headend 1410 provide speech processing capabilities. Augmented Headend 1410 may be used to handle speech processing for nodes 120 and 124. Augmented metropolitan Headend 1410 may be used to handle speech processing for Headends 100 and 106. Headends 100 and 106, and nodes 120 and 124 are not augmented by this invention.

Alternatively, augmented Headend 1410 may control the optimized upstream communication as disclosed in the co-pending application Ser. No. 09/679,115, entitled "Increased Bandwidth in Aloha-based Frequency Hopping Transmission Systems". Node 1310 may support the optimized upstream communication. Upstream communication from augmented node 1310 and from augmented Headend 1410 may employ the same upstream communications protocol, or different upstream communications protocol.

Note that metropolitan Headend 1410 may not be involved in the optimized upstream communication scheme.

Alternatively, metropolitan Headend 1410 may be involved in the optimized upstream communication scheme. Augmented metropolitan Headend 1410 may further control the optimized upstream communication as disclosed in the co-pending application Ser. No. 09/679,115, entitled "Increased Bandwidth in Aloha-based Frequency Hopping Transmission Systems". Augmented Headend 1410 may support the optimized upstream communication between node 1310 and set-top apparatus 1100-180. Node 1310 may then carry out the optimized upstream communication. Upstream communication from augmented node 1310 and from augmented Headend 1410 may employ the same upstream communications protocol, or different upstream communication protocols.

Note that in certain network installations, there may be no metropolitan Headend, augmented or otherwise. Further, in certain network installations, there may only be one Headend.

FIG. 22 depicts a remote control unit 1000-180 coupled 1002-180 to set-top apparatus 1100-180, communicating via a two-stage wireline communications system containing a wireline physical transport 1200 to a distributor node 126. Distributor node 126 interfaces to a wireline communications loop including an augmented Headend 1414. The network may further support a communications loop including augmented metropolitan Headend 1410, in accordance with the invention.

Augmented Headend 1414 may control and support optimized upstream communication as disclosed in the co-pending application Ser. No. 09/679,115, entitled "Increased Bandwidth in Aloha-based Frequency Hopping Transmission Systems" by Calderone and Foster, both inventors of this application and commonly assigned to AgileTV, incorporated herein by reference.

Augmented Headend 1414 may provide cached content such as popular videos for a VOD service. Speech processing may be performed in many situations at augmented Headend 1414

Augmented metropolitan Headend 1410 may further provide speech processing capabilities. Augmented metropolitan Headend 1410 may be used to handle speech processing requirements for Headends 100 and 106, which are not augmented by this invention.

The couplings between STB 1100-180, Node 126 and Headend 1410 may also support bi-directional communication allowing the STB 1100-180 to receive multiple television channels and allowing STB 1100-180 to signal at least limited information to augmented Headend 1410, which may well include management of Pay-per-View and other services.

Alternatively, augmented metropolitan Headend 1410 may control the optimized upstream communication as disclosed in the co-pending application Ser. No. 09/679,115, entitled Increased Bandwidth in Aloha-based Frequency Hopping Transmission Systems. Augmented Headend 1414 may then support the optimized upstream communication. Upstream communication from augmented node 126 and

from augmented Headend **1410** may employ the same upstream communications protocol, or a different upstream communications protocol.

Note that metropolitan Headend **1410** may not be involved in the optimized upstream communication scheme.

Alternatively, metropolitan Headend **1410** may be involved in the optimized upstream communication scheme. Augmented metropolitan Headend **1410** may further control the optimized upstream communication as disclosed in the co-pending application Ser. No. 09/679,115, entitled Increased Bandwidth in Aloha-based Frequency Hopping Transmission Systems. Augmented Headend **1410** may support the optimized upstream communication between node **126** and set-top apparatus **1100-180**. Node **126** may then carry out the optimized upstream communication. Upstream communication from node **126** and from augmented Headend **1410** may employ the same upstream communications protocol, or different upstream communication protocols.

Note that in certain network installations, there may be no metropolitan Headend, augmented or otherwise. Further, in certain network installations, there may only be one Headend.

FIG. **23** depicts a detail block diagram of an augmented distributor node **1310**, coupled to wireline physical transport **1200** and coupled to the wireline communications loop of FIG. **21**.

Note that there may be one transport **1200**. Transceiver **1320** couples to transport **1200** to provide uplink **1204** and downlink **1202** communications between STB's **1100** of FIG. **21**.

There may be multiple transports **1200** corresponding to wireline circuits servicing combinations of at least one of the interfaces **170**, **174** and **178**, as shown in FIGS. **1**, **2**, **21** and **22**. Transceiver **1320** couples to transport **1200** to provide uplink **1204** and downlink **1202** communications between STB's **1100** of FIGS. **21** and **22**. Note that transceiver **1320** may provide multiple uplinks **1204**. Transceiver **1320** may provide multiple downlinks **1202**.

Transceiver **1320** may include transmitter **1324** providing downlink **1202** communications to wireline physical transport **1200**. Multiple downlinks **1202** may be provided by transmitter **1324**. Alternatively, multiple downlinks **1202** may be provided by more than one transmitter **1324**.

Transceiver **1320** may include receiver **1322** providing uplink **1204** communications from wireline physical transport **1200**. Multiple uplinks **1204** may be provided by receiver **1322**. Alternatively, multiple uplinks **1204** may be provided by more than one receiver **1322**.

Wireline physical transport **1200** may further include separate uplink physical transport **1204** and downlink physical transport **1202**. Various embodiments of the invention may include multiple uplink physical transports **1204**. Various embodiments of the invention may include multiple downlink physical transports **1202**. There may be a difference in the number of uplink physical transports **1204** and the number of downlink physical transports **1202**.

Note that there may be a difference in the number of transmitters **1324** and the number of receivers **1322**.

By way of example, referring to FIG. **21**, one downlink stream **1202** may be sent to interface **170**, and a second downlink stream, not shown in FIG. **21** or FIG. **23**, may be sent to interface **174**. One uplink stream **1204** may be received from interface **170**, and a second uplink stream, not shown in FIG. **21** or FIG. **23**, may be received from interface **174**.

Hereafter, the discussion focuses on a single transmitter **1324** providing as many downlinks **1202** and on a single receiver **1322** providing as many uplinks **1204** as required. This is done strictly to simplify the discussion and is not meant to imply any limitation on the invention.

Uplink **1204** communication includes a back channel. This back channel includes multiple identified speech channels from multiple user sites (STBs) **1100**, as shown in FIGS. **21** and **22**. Receiver **1322** provides **1326** a back channel to speech engine **1330**. Speech engine **1330** performs at least the operations of FIG. **10**.

Speech engine **1330** further communicatively interacts **1374** with network interface **1370**. Network interface **1370** couples **1372** to other network components. Network coupling **1372** may further include a predominantly input coupling **1306** and a predominantly output coupling **1308**.

Network interface **1370** may communicatively interact **1376** with local system management and billing **1380**. Local system management and billing **1380** may include systems management and billing data relevant to the elements of the overall network managed or controlled through the local node.

Local system management and billing **1380** may further maintain systems management and billing data relevant to the elements of the overall network managed or controlled through the local node.

Local system management and billing **1380** may include a cache of systems management and billing data relevant to the elements of the overall network managed or controlled through the local node.

Receiver **1322** provides **1326** a back channel to speech engine **1330**. Speech engine **1330** performs at least the operations of FIG. **10**. Speech engine **1330** further communicatively interacts **1374** with network interface **1370**.

Speech engine **1330** may communicatively interact **1382** with system management and billing **1380**. Speech engine **1330** may send **1382** system management and billing **1380** requests for user account information, including, but not limited to, credit information, authorization profiles for channel viewing, credit limits, and parental safeguards. Speech engine **1330** may send **1382** system management and billing **1380** user passwords, confirmation of commitments, commands regarding authorization levels of other users, such as children within a user's household.

Speech engine **1330** may receive **1382** from system management and billing **1380** credit information, authorization profiles for channel viewing, credit limits and parental safeguards. Speech engine **1330** may receive **1382** from system management and billing **1380** confirmation of acceptance of contracts from financial engines. Acceptance of contracts from external financial engines may be received in a number of different ways, including through interactions with network interface **1370** or other external communications networks **1312**.

System management and billing **1380** may further include interfaces to other external communications networks. Such circuitry is not the subject of this invention and is not discussed further herein.

Similarly, the mechanism of contract acceptance between system management and billing **1380** and financial engines is not the subject of this invention and is not discussed further herein.

Speech engine **1330** may receive **1338** from content engine **1340** content status information. Content status information may include a list of available services through the coupled network(s). Content status information may further include a list of services available within the node.

Content status information may include a list of content items cached through the network. Content status informa-

tion may further include a list of content items cached at the node. Cached content items may include Interactive Program Guide (IPG) listings for some period of time. Cached content items may include one or more video sequences provided through Video On Demand (VOD) or Pay-Per-View services.

Content engine **1340** communicatively interacts **1378** with network interface **1370**. Content engine **1340** may provide at least one locally generated multi-media stream **1342** to Multiplexor engine **1360** as well as at least one multi-media stream **1344** received **1378** through network interface **1370**. Content engine **1340** may modify multi-media stream **1344** received **1378** from network interface **1370** through network input **1306**. Content engine **1340** may stimulate **1378** network interface **1370**, altering network output **1308** from network interface **1370**.

Content engine **1340** may have more than one registered application server each presenting static menu content and dynamic content. A speech processor within speech engine **1330** may be presented a grammar based upon the static and dynamic content of multiple registered application servers. The merged services may be presented to the users as specialized, multi-application service menus provided by the speech engine **1330** to the users.

Speech engine **1330** may generate one or more channels of speech response content through coupling **1332** to modulator engine **1350**. Modulator engine **1350** may further provide **1334** status and reliability information to speech engine **1330**. Speech response content channels presented through coupling **1332** may be digital. Speech response content channels may be presented as bits or clusters of bits of a specific bit width.

Multiple channels may be multiplexed onto coupling **1332** by speech engine **1330**. The multiplexing mechanism onto channel **1332** may perform time division multiplexing. Modulator engine **1350** may demultiplex multiple multiplexed channels received through coupling **1332**. Modulator engine **1250** may convert one or more demultiplexed channels into modulated channels, or modulated collections of channels, presented **1352** and **1354** to multiplexor engine **1360**.

Multiplexor engine **1360** accepts the multiple locally generated channels **1352**, **1354** and **1342** as well as the locally received, and possibly modified, external stream **1344** to create at least one merged stream **1362**. Multiplexor engine **1360** may create more than one merged stream, e.g. **1364**.

Transmitter **1324** receives at least one merged stream **1362** from multiplexor engine **1360** to generate at least one downlink stream **1202** of physical transport **1200**. Transmitter **1324** may receive more than one merged stream **1364** from multiplexor engine **1360**.

Speech engine **1330** may further interact **1312** with an external network. Such an interaction may involve at least one wireline physical transport layer. The wireline physical layer may support at least one or a combination of communication protocols using optical, infra-red and radio frequency regions of the electromagnetic spectrum. Network interactions **1312** may support message passing protocols, including, but not limited to, TCP-IP. Network interactions **1312** may further support communications with the Internet and World Wide Web.

FIG. **24** depicts an alternative detail block diagram of an augmented distributor node **1310**, coupled to wireline physical transport **1200** and coupled to the wireline communications loop of FIG. **21**.

Receiver **1322** provides a back channel to speech engine **1330** through interactions **1328** with switch **1390** delivering **1392** the back channel to speech engine **1330**.

Speech engine **1330** communicatively interacts **1374** with network interface **1370** by the speech engine **1330** interacting **1392** with switch **1390** which interacts **1398** with network interface **1370**.

Network interface **1370** may communicatively interact with local system management and billing **1380** by communicatively interacting **1398** with switch **1390**, which communicatively interacts **1396** with system management and billing **1380**.

Speech engine **1330** may communicatively interact **1382** with system management and billing **1380** by communicatively interacting **1392** with switch **1390**, which communicatively interacts with system management and billing **1380**.

Speech engine **1330** may receive content status information from content engine **1340** by content engine **1340** interacting **1394** with switch **1390**, which delivers **1392** the content status information to speech engine **1330**.

Content engine **1340** communicatively interacts with network interface **1370** by content engine **1340** communicatively interacting with switch **1390** and network interface **1370** communicatively interacting with switch **1390**.

Switch **1390** may support digital interfaces. Switch **1390** may include a circuit switch. The circuit switch may support Ethernet protocols. Switch **1390** may include an ATM switch. Switch **1390** may support analog interfaces. Such analog interfaces may include wavelength division multiplexing. Switch **1390** may be composed of more than one switch.

The invention may include various combinations of direct interconnections and switch networks as shown in FIGS. **23** and **24**.

FIG. **25** depicts a generic block diagram of a prior art Headend **104** as shown in FIG. **3**.

Uplink **138** communication includes a back channel.

This back channel includes multiple response channels from multiple user sites STBs, as shown in FIGS. **1** and **2**. Receiver **1422** provides **1427** a back channel to content engine **1440**.

Network interface **1470** may communicatively interacts **1476** with Headend system management and billing **1480**. Headend system management and billing **1480** may include systems management and billing data relevant to the elements of the overall network managed or controlled through the Headend.

Headend system management and billing **1480** may further maintain systems management and billing data relevant to the elements of the overall network managed or controlled through the Headend.

Headend system management and billing **1480** may include a cache of systems management and billing data relevant to the elements of the overall network managed or controlled through the Headend.

System management and billing **1480** may further include interfaces to other external communications networks. Such circuitry is not the subject of this invention and is not discussed further herein.

Content status information may include a list of content items cached through the network. Content status information may further include a list of content items cached at the node. Cached content items may include Interactive Program Guide (IPG) listings for some period of time. Cached content items may include one or more video sequences provided through Video On Demand (VOD) or Pay-Per-View services.