

PROVISIONAL APPLICATION COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION under 37 CFR 1.53(b)(2).

A/PROV

11/17/99


		Docket Number	SAR 13807P	Type a plus sign (+) inside this box	+
INVENTOR(S)/APPLICANT(S)					
LAST NAME	FIRST NAME	MIDDLE INITIAL	RESIDENCE (CITY AND EITHER STATE OR FOREIGN COUNTRY)		
Murdock	Michael		East Windsor, NJ		
Pearson	John		Lawrenceville, NJ		
Sajda	Paul		Princeton, NJ		
TITLE OF THE INVENTION (280 characters max.)					
A SYSTEM AND METHOD FOR VOICE NAVIGATION FOR NETWORKED TELEVISION-CENTRIC INFORMATION SERVICES WITH BACK CHANNEL					
CORRESPONDENCE ADDRESS					
Sarnoff Corporation Director, Law and Patents 201 Washington Road CN 5300 Princeton					
STATE	New Jersey	ZIP CODE	08543-5300	COUNTRY	US
ENCLOSED APPLICATION PARTS (check all that apply)					
<input checked="" type="checkbox"/> Specification	Number of Pages	5	<input type="checkbox"/> Small Entity Statement		
<input checked="" type="checkbox"/> Drawing(s)	Number of Sheets	1	<input type="checkbox"/> Other (specify)		
METHOD OF PAYMENT (check one)					
<input type="checkbox"/> A check or money order is enclosed to cover the Provisional filing fees			PROVISIONAL FILING FEE AMOUNT(S)		\$150.00
<input checked="" type="checkbox"/> The Commissioner is hereby authorized to charge filing fees and credit Deposit Account Number: 20-0782					

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.

☒ No.

☐ Yes, the name of the U.S. Government agency and the Government contract number are: _____

Respectfully Submitted,

SIGNATURE 
 TYPE or
 PRINT NAME Kin-Wah Tong

Date: 11/17/99

REGISTRATION NO. 39.400

☐ Additional inventors are being named on separately numbered sheets attached hereto.

EXPRESS MAIL CERTIFICATION

"Express Mail" mailing label number EL446266677US

Date of deposit 11/17/99

I hereby certify that this paper or fee is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated above and is addressed to the Assistant Commissioner for Patents, Box Provisional Patent Application, Washington, D.C. 20231.

Susan McCormack

Signature of person mailing paper or fee

Susan McCormack

Name of person mailing paper or fee

A System and Method for Voice Navigation for Networked Television-Centric Information Services with Back Channel

The present invention relates to an apparatus, system and
5 concomitant method for voice control of a networked television-centric information service that utilizes a backward channel capability.

DETAILED DESCRIPTION

While speech-controlled consumer appliances are attractive for
10 many reasons, accurate and reliable speech recognition requires a fairly powerful computer. It is the first advantage of this invention to provide a central speech recognition server that can be used by many speech-controlled appliances that are connected by a network. This means that a speech-controlled appliance can be delivered in a much more economical
15 manner than if it required a fast CPU with a large amount of memory. It is the second advantage of this invention to provide a mechanism, the network's backward channel, for transmitting voice control data to the network information source. The central speech recognition server is contained at the network information source. It processes speech control
20 data arriving through the backward channel and then sends the appropriate programming modifications through the forward channel as requested by the viewer. One of the principal reasons for speech recognition inaccuracies is the background room noise that masks the speech signal. Even a small amount of noise from other speakers can lead
25 to a large number of speech errors. Additionally, noise in the speech signal can make it very difficult to find an economical code for transmission. It is the third advantage of this invention to provide local speech enhancement in order to achieve an optimal coding for economical transmission through the backward channel. Since multiple microphones
30 can be used to accurately focus on the viewer's speech, it is the fourth advantage of this invention to use multiple microphones to acquire the speech signal and then use a local signal separation method in the home entertainment processor to extract the clean signal prior to speech coding.

System Description

The following is a brief definition of each of the important terms in this invention. The *Home Entertainment Processor (HEP)* is the core component of a home entertainment system which is primarily a

5 television viewing system. The HEP is an electronic device for receiving network programming content through the forward channel and transmitting speech control data through the backward channel. The *forward channel* is the channel in which the programming content from the information source is transmitted into the HEP. The back channel is
10 the channel in which the Voice Control Data from the HEP is transmitted to the information source. In this invention, the HEP is controlled by a speech recognition engine that resides in the Information Source. The content from the *information source* could be web content, video-on-demand, cable television, or a local video cassette player.

15 A *Hand-Held Controller (HHC)* is used by the viewer in a manner that is similar to the way current television remote controls are used with the exception that the viewer makes selections by speaking into the HHC rather than pushing buttons. The viewer's speech command ("*Show me
20 all of the NCAA football games on this afternoon*") is acquired from the HHC with one or more microphones, which is then transmitted to the HEP. It is also in the current embodiment to acquire a video signal of the speaker using a video camera in the Hand-Held Controller. This would also be transmitted to the HEP and processed in conjunction with the
25 speech signal. The speech signal is then used to determine the speaker's identification. The speech signal is enhanced to reduce noise in a manner that improves the speech recognition accuracy and the efficiency of the speech coder. Finally, the speech signal is coded for efficient transport over the back channel and sent out to the Back Channel Multiplexor.

30 It should be noted that the HEP could also be connected to the *Home Compute Server*, which is a computer device for controlling non-entertainment compute devices in the home (e.g. communications, security, personal computing, kitchen appliances, etc.). Figure 1 illustrates a block diagram of an illustrative embodiment of the present

invention. Each of the components is described in more detail in the following section.

Component Descriptions.

- 5 The functionality of each of the components of this invention in one embodiment is as follows:

- Hand-Held Controller - This component is a small electronic device that is held by the viewer and used to navigate through the programming
10 content. It contains one or more microphones, a video camera and a radio-frequency transmitter for sending the speech and video data to the HEP. The current embodiment of this invention contains push button controls for viewers that do not wish to use speech control.
- 15 Home Entertainment Processor - This component is an electronic device for receiving network content through the network's forward channel. It contains software modules that perform speaker identification, speech enhancement and coding and video tracking. The video component of the system is used to extract visual information (such as lip location) from the
20 viewer that will render the speech processing more accurate. The speaker identification is used by the speech server to perform a more accurate recognition by employing a user profile that contains a statistical model of an individual's preferences and command patterns. The coded speech commands are sent out to the network's backward channel in order that
25 the speech commands can be processed by a speech recognition engine at the information source.

- Backward Channel Multiplexor - This component mixes the backward channel content received from several HEPs into one signal that is sent to
30 the Information source. In situations in which a large number of viewers are likely to speaking simultaneously, the BCM can be used to delay and stagger the content, thus avoiding channel overload.

Information Source - This component is located at the site of the information provider. The information could be web content, video-on-demand, or cable television. However, in any case, this component contains the central speech recognition server. The speech recognition server is connected to electronic programming guides, vocabulary lists and other meta-data. Also, an inference engine for processing user profile information is used by the Information Source to help guide the speech recognition engine.

10 Conclusion

It is the first advantage of this invention to provide a central speech recognition server that can be used by many speech-controlled appliances that are connected by a network. This means that a speech-controlled appliance can be delivered in a much more economical manner than if it required a fast CPU with a large amount of memory.

It is the second advantage of this invention to provide a mechanism, the network's backward channel, for transmitting voice control data to the network information source. The central speech recognition server is contained at the network information source. It processes speech control data arriving through the backward channel and then sends the appropriate programming modifications through the forward channel as requested by the viewer. One of the principal reasons for speech recognition inaccuracies is the background room noise that masks the speech signal. Even a small amount of noise from other speakers can lead to a large number of speech errors. Additionally, noise in the speech signal can make it very difficult to find an economical code for transmission.

It is the third advantage of this invention to provide local speech enhancement in order to achieve an optimal coding for economical transmission through the backward channel. The speech enhancement also eliminates noise from the speech signal so that the recognition is much more accurate.

It is the fourth advantage of this invention to use multiple microphones to acquire the speech signal. Multiple microphones can be

used to accurately focus a beam on the viewer's speech. A local signal separation method in the home entertainment processor is used to extract the clean signal prior to speech coding.

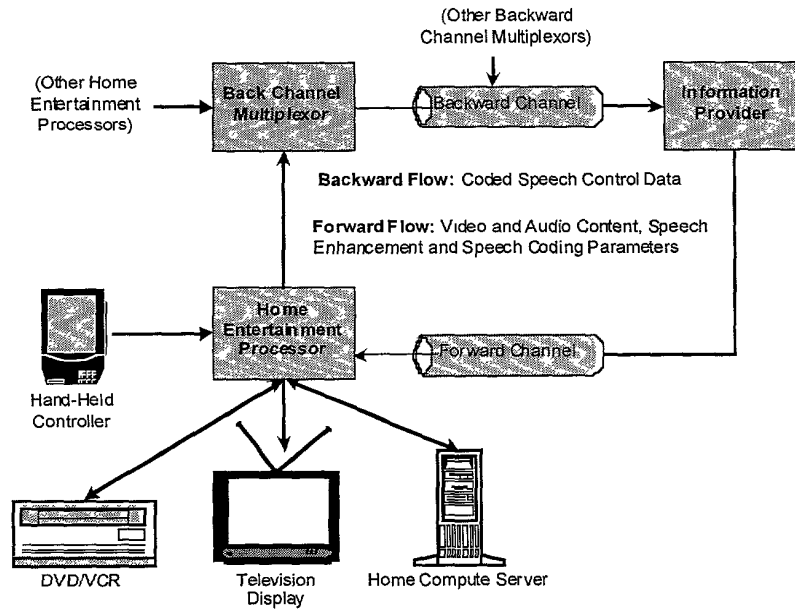


Figure 1. Block Diagram of System.