



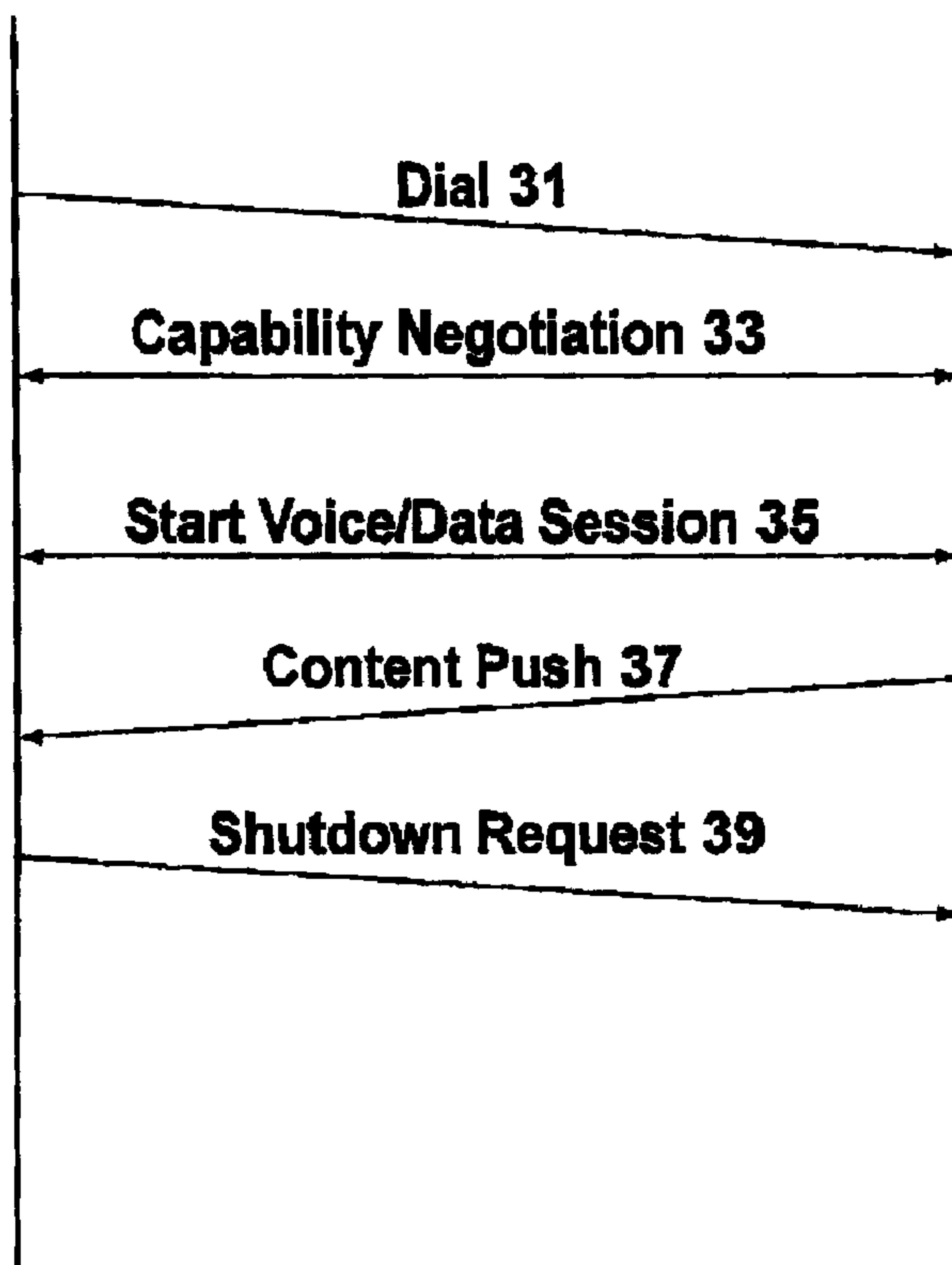
(22) Date de dépôt/Filing Date: 2007/06/28
 (41) Mise à la disp. pub./Open to Public Insp.: 2007/12/29
 (45) Date de délivrance/Issue Date: 2012/10/23
 (30) Priorité/Priority: 2006/06/29 (EP06116300.2)

(51) Cl.Int./Int.Cl. *H04L 12/16* (2006.01),
H04L 12/18 (2006.01), *H04W 4/16* (2009.01),
H04W 88/06 (2009.01)
 (72) Inventeur/Inventor:
 LAI, FREDERICK CHEE-KIONG, CA
 (73) Propriétaire/Owner:
 RESEARCH IN MOTION LIMITED, CA
 (74) Agent: PERRY + CURRIER

(54) Titre : TELEPHONE-FURETEUR HYBRIDE PSEUDO-ENRICHI
 (54) Title: PSEUDO-RICH HYBRID PHONE/BROWSER

ENDPOINT 10

ENDPOINT 12



(57) **Abrégé/Abstract:**

A markup language specification is set forth for providing pseudo-rich media during phone calls, and to implement two endpoints that support this specification. Each implemented endpoint functions as a half-phone and half-browser, where the phone call consists partly of the traditional full-duplex audio stream between callers, supplemented by pseudo-rich media being transmitted from one party to the other. The pseudo-rich media includes, but is not limited to, text, pictures and hyperlinks.

ABSTRACT

5 A markup language specification is set forth for providing pseudo-rich media
during phone calls, and to implement two endpoints that support this specification. Each
implemented endpoint functions as a half-phone and half-browser, where the phone call
consists partly of the traditional full-duplex audio stream between callers, supplemented
by pseudo-rich media being transmitted from one party to the other. The pseudo-rich
10 media includes, but is not limited to, text, pictures and hyperlinks.

Pseudo-rich Hybrid Phone/Browser

[0001] The following is directed in general to communication devices, and more particularly to a hybrid phone/browser for providing simultaneous audio and visual content while consuming minimal bandwidth.

[0002] Phone applications that use cellular networks or WLAN networks are traditionally considered to be audio applications. The content of a traditional phone call is typically limited to a full duplex audio stream that is shared between two or more callers. One problem with audio-only connections is that information is shared very slowly, and is limited by the ability of the listening party to hear the talking party. Some types of information, such as phone numbers, product ID numbers, menu selections, etc., are not well communicated through audio. Background noise, drops in voice quality and the time required to hear an entire pre-recorded audio stream make an indication of specific information unduly laborious and grueling.

[0003] Videoconferencing applications have attempted to solve the limitations of audio-only communications by allowing users to send video streams to each other during a call, where the video is captured by respective video cameras (or other video streaming mechanisms) in order to convey images of each caller. The video streams are then transmitted between communication peers for rendering in real-time.

[0004] One significant disadvantage of videoconferencing applications is that the bandwidth consumed is extremely large while the information presented is limited only to an image of the remote peer (i.e. the information does not provide much in terms of value added).

[0005] It is also known in the art to provide a cellular-phone with a Web browser. However, there is no integration between the phone and browser applications in such prior art devices.

[0006] According to an aspect of this specification, simultaneous audio and visual content are preferably provided at low bandwidth, to provide dynamic, useful information to the user. A markup language specification is set forth for providing pseudo-rich media during phone calls and for implementing two endpoints that support this specification. Each implemented endpoint preferably functions as a half-phone, half-browser (or half-server, as the case may be). In other words, a phone call consists partly of the traditional full-duplex audio stream between the parties and is supplemented by pseudo-rich media being transmitted from one of the parties to the other. It is contemplated that the pseudo-rich media include, but not be limited to, text, pictures and hyperlinks.

Brief Description of the Drawings

[0007] The foregoing will be better understood with reference to the description and to the following drawings, in which:

5 **[0008]** Figure 1, including Figures 1A, 1B, 1C and 1D, is a schematic representation of a mobile device with a user interface supporting communication via the specification set forth herein;

[0009] Figure 2 is a block diagram showing connection of the mobile device of Figure 1 with a server for providing interactive voice response (IVR);

10 **[0010]** Figure 3 is a simplified sequence diagram showing exemplary communication between the mobile device and the server of Figure 2; and

[0011] Figures 4A, 4B and 4C are internal architecture diagrams for implementing various exemplary embodiments of the user interface for the mobile device of Figure 2.

Description of Preferred Embodiments

15 **[0012]** With reference to Figures 1 and 2, a first user endpoint is connected to a second user endpoint over a peer-to-peer network. More particularly, a mobile device 10 (first endpoint) having a pseudo-rich phone browser, is connected through a proxy, a gateway or a firewall (designated generally by 11A) to the network 14. It will be appreciated that this connection can include a wireless connection, for a cellular phone,
20 for example. The mobile device 10 includes a microphone 13, speaker or earpiece 14 and a display 15.

[0013] A server 12 (second endpoint) is connected to the network 14 via, for example, a proxy, a gateway, a firewall or a load balancer (designated generally by 11B). The server can, for example, include an interactive voice response system (IVR). The network
25 14 supports a pseudo-rich communication specification, as further discussed below.

[0014] According to the example of Figures 1A, 1B, 1C and 1D, the user of mobile device 10 places a call to the ABC Company customer support helpline, which utilizes an IVR server 12 that supports the pseudo-rich specification set forth herein.

30 **[0015]** Once the call between device 10 and server 12 has been established, an automated voice response from the IVR greets the user with an audio message that is reproduced via the speaker 14 at device 10, such as: "Welcome to the ABC Company consumer helpline... etc.". At the same time, through the markup language (i.e. script) discussed below, text corresponding to the voice announcement is displayed as an image at display 15, via the phone browser application (Figure 1A). The text may be
35 accompanied by a background picture of the company logo or other suitable images. As the script continues, it asks "for service in English, press 1, pour le service en français

appuyer sur le 2. To hear this information again, press star". At the same time, markup information is pushed to the phone at endpoint 10 (Figure 1B) to display: "Press: 1 for English, 2 pour le français". In response, the user can, optionally, press "*" to hear the information again from the automated attendant. Since the phone supports pseudo-rich media, however, the user can merely glance at the screen of the phone to view the information rather than pressing "*" to hear the information again.

[0016] Alternatively, if the server 12 incorporates voice recognition technology then the user may respond by issuing voice commands that are recognized by the server 12 and then acted upon. Such voice recognition systems are well known in the art.

10 **[0017]** During the call, the phone 10 receives messages from the IVR server 12 out of band with the audio connection. That is, the user at phone 10 does not hear the data being transmitted to the phone, while the phone decodes the data for display.

[0018] The user can continue navigating through the IVR system to find the address of the organization. As the IVR reads out the information for the user to hear, the information is simultaneously displayed, as shown in Figure 1C.

15 **[0019]** After receiving the desired information, the user requests shutdown by, for example, responding "no" to the question "Do you require any further information?" (Figure 1D). In response to receipt of the shutdown request, the call is ended, while retaining the graphic information concerning a contact address on the display screen of the phone 10.

[0020] Figure 3 shows a simplified sequence diagram of messages exchanged to provide simultaneous audio and visual communication between the mobile device 10 and the server 12, according to an exemplary embodiment. The user of mobile device 10 begins by dialing the appropriate number to connect with the second endpoint (Dial 31). After establishing a connection, the pseudo rich phone browser within device 10 and the IVR server 12 negotiate capabilities (Capability Negotiation 33). When the capabilities of the pseudo rich phone browser are determined by the IVR, the voice and data session is started (Start Voice/Data Session 35). The IVR server 12 sends audio to the phone 10 while carrying out speech recognition as well as DTMF tone detection on audio received from the phone. Data content and audio are sent simultaneously by the IVR server 12 to the phone 10 based on audio responses received from the phone (Content Push 37). In carrying out this communication, packet-switched data is transmitted from the IVR server 12 to the phone 10. Data can be pushed to the phone any number of times. In response to receipt of the shutdown request (Shutdown Request 39), the call is ended.

35 **[0021]** The IVR example of Figures 1 – 3 is but one of many possible examples of a method and apparatus for providing simultaneous full-duplex audio and a pseudo-rich

media stream between parties to a call. Additional examples include creating a custom "voice page" on a home server, similar to well-known individual home pages, but which is accessible via a browser-enabled phone 10, and provisioning of a desktop phone browser, as discussed in greater detail below with reference to Figure 5.

5 **[0022]** Figure 4A shows an internal architecture for implementing the user interface 40 within device 10 of Figure 1, according to one embodiment. According to this embodiment, separate browser and phone applications 41 and 43 are employed while the server 12 coordinates timing for pushing the pseudo-rich browser data, audio and speech recognition. The browser and phone components represent the highest layer (Application
10 Layer 7) of the Open Systems Integration (OSI) model of data networking. Data protocol layer 44 and phone signaling/audio protocol 45 form the Presentation Layer of the OSI model. Transport protocol stacks 47A and 47B (OSI Layer 4) manage end-to-end control and error checking to ensure complete data transfer. Packet data stack 49 forms the data link layer (Layer 2) for node-to-node validity and integrity of the data transmission.
15 Hardware 51 is the physical layer (Layer 1) of the OSI model responsible for passing bits onto and receiving them from the connecting medium.

[0023] The data structure of the packets that are transmitted is based on a modified version of the Voice Extensible Markup Language (VoiceXML). The IVR script is written to allow synchronization of voice and data for playback and display. As described above,
20 images are displayed while sounds are simultaneously played back. Exemplary Voice XML code for implementing the pseudo-rich hybrid phone browser of the present application is as follows:

```
<vxml version="2.0" xmlns="http://www.w3.org/2001/vxml"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
25   xsi:schemaLocation="http://www.w3.org/2001/vxml
  http://www.w3.org/TR/voicexml20/vxml.xsd">

  <!--begin editable region-->
  <table width="100%" border="0" cellspacing="4" cellpadding="4">
30   <tr align="left" valign="top">
     <td width="46%">
     <br />
     <span class="cm">You have reached ABC Company.
35   Please say the extension or name of the person you wish to reach.</p>

  <!--INSERT some cool graphics code here: animated icon, interesting
     visual effects, etc.-->
```

```

<table width="590" height="221"
      background="/images/home/8700_7100_ABC_home.jpg"
      style="background-repeat:no-repeat">
<tr>
5  <td width="95" height="150"><a href="http://www.ABC.com/products
      /index.shtml" target="other"></a></td>
<td width="495" height="221" rowspan="2">
      <a href="http://www.ABC.com/products/ index.shtml"
10  target="other"></a></td>
</tr>
<tr>
<!--<td width="95"><a href="http://www.ABC.com/news.shtml"
15  target="other"></a></td-->
<td></td>
</tr>
20 </table>
</td>
</tr>
</table>
<!--end editable region-->
25
<form id="no_bargein_form">
<property name="bargein" value="false"/>
<block>
<prompt>
30  This introductory prompt cannot be barged into.
</prompt>
<prompt>
      And neither can this prompt.
</prompt>
35 <prompt bargein="true">
      Thanks for calling ABC! Do you know the extension of the
      person you wish to reach?
</prompt>
</block>
40
<field type="boolean">
<prompt>
      Please say yes or no.
</prompt>
45 </field>

```

<!--more prompts and voice recognition code and more text displayed on the screen.-->

</form>

5 </vxml>

[0024] Turning to Figures 4B and 4C alternative internal architectures are depicted for implementing the user interface of Figures 1 - 3. Referring first to Figure 4B, an embodiment is illustrated in which a video application 55 feeds the images, rather than a browser application as
10 in Figure 4A. The video and phone applications 55 and 43, are separate as in the architecture of Figure 4A. The server 12, however, coordinates timing for pushing video images, sound and when to carry out speech recognition.

[0025] Referring to Figure 4C, the video and audio are integrated in the same application 59 and protocol 61 as in, for example, a videophone. Server 12 (in this case a video server)
15 therefore coordinates timing for pushing video images, sound and when to carry out speech recognition based on state.

[0026] A person skilled in the art, having read this description, may conceive of variations and alternative embodiments. For example, the data structure of the packets that are transmitted is not limited to a modified version of VoiceXML as other data structures and protocols are
20 possible. It is contemplated that HTML content could be pushed from the IVR to the first endpoint by embedding an HTML page in the payload section of a Session Initiation Protocol (SIP) message (RFC3261 by Rosenberg, J. et al, entitled SIP: Session Initiation Protocol, copyright The Internet Society, June 2002). A SIP INFO method (RFC2976 by Donovan, S., entitled SIP: Info Method, copyright The Internet Society, October 2000), or another similar
25 method, can be employed. It is also contemplated that other media and audio/video sequencing protocols can be employed. For example, an audio/video protocol that is similar to Macromedia Flash™ can be used while routing voice traffic on the audio end, as well as speech recognition. Still other variations and modifications may occur to those skilled in the art.

[0027] All such variations and alternative embodiments are believed to be within the ambit of
30 the claims appended hereto.

Claims:

1. A method of providing simultaneous audio and visual content for a portable electronic device in communication with an Interactive Voice Response server comprising:
- 5 transmitting from said Interactive Voice Response (IVR) server to said portable electronic device audio content over a full duplex audio link and pseudo-rich media content relating to and synchronized with said audio content over a data link;
- 10 receiving messages from said portable electronic device at said Interactive Voice Response (IVR) server responsive to said audio content and pseudo-rich media content; and
- initiating generation of further audio content and pseudo-rich media content within said Interactive Voice Response (IVR) server for transmission to said portable electronic device responsive to said messages.
- 15 2. The method of claim 1, wherein said pseudo-rich media content conforms to a markup language specification.
3. The method of claim 1 or 2, wherein said pseudo-rich media content comprises at least one of text, pictures and hyperlinks.
- 20 4. The method of claim 1, 2 or 3 comprising reproducing said audio content from a speaker of said portable electronic device; and reproducing said pseudo-rich media content on a screen of said portable electronic device.
- 25 5. A communication system, comprising:
- an Interactive Voice Response server for generating synchronized audio and pseudo-rich media content relating to and synchronized with said audio content;
- means for transmitting the audio content over a full duplex audio link and the pseudo-rich media content over a data link, and
- 30 a device for receiving and reproducing said synchronized audio and pseudo-rich media content, and for transmitting messages to said server responsive to said audio content and pseudo-rich media content, whereupon said server generates further audio content and pseudo-rich media content responsive to said messages for sending to said device.

6. The communication system of claim 5, wherein said device comprises separate browser and phone applications, separate data and phone signaling/audio protocols, separate transport protocol stacks, a packet data stack and a physical layer.
- 5 7. The communication system of claim 5, wherein said device comprises separate video and phone applications, separate video and phone signaling/audio protocols, separate transport protocol stacks, a packet data stack and a physical layer.
- 10 8. The communication system of claim 5, wherein said device comprises an integrated video and audio application an integrated video and audio protocol, a transport protocol stack, a packet data stack and a physical layer.
9. The communication system of any of claims 5 to 8, wherein said pseudo-rich media content conforms to a markup language specification.
- 15 10. The communication system of claim 9, wherein said pseudo-rich media content comprises at least one of text, pictures and hyperlinks.
11. The communication system of any of claims 5 to 10, wherein said server implements
20 an Interactive Voice Response (IVR) system comprising a voice recognition capability for recognizing and responding to user voiced commands.
12. An Interactive Voice Response (IVR) server for providing simultaneous audio and visual content to a portable electronic device, comprising:
- 25 means for transmitting to said portable electronic device audio content and pseudo-rich media content relating and synchronized to said audio content, such that said audio content and pseudo-rich media content are reproduced within said portable electronic device;
- means for receiving messages from said portable electronic device responsive to user input; and
- 30 means for initiating generation of further audio content and pseudo-rich media content and transmitting further audio content and pseudo-rich media content said to said portable electronic device responsive to receipt of said messages.

13. The Interactive Voice Response (IVR) server of claim 12, wherein said pseudo-rich media content conforms to a markup language specification.

14. The Interactive Voice Response (IVR) server of claim 12 or 13, wherein said pseudo-rich media content comprises at least one of text, pictures and hyperlinks.

15. The Interactive Voice Response (IVR) server of claim 12, 13 or 14, wherein said server implements an Interactive Voice Response (IVR) system comprising a voice recognition capability for recognizing and responding to user voiced commands.

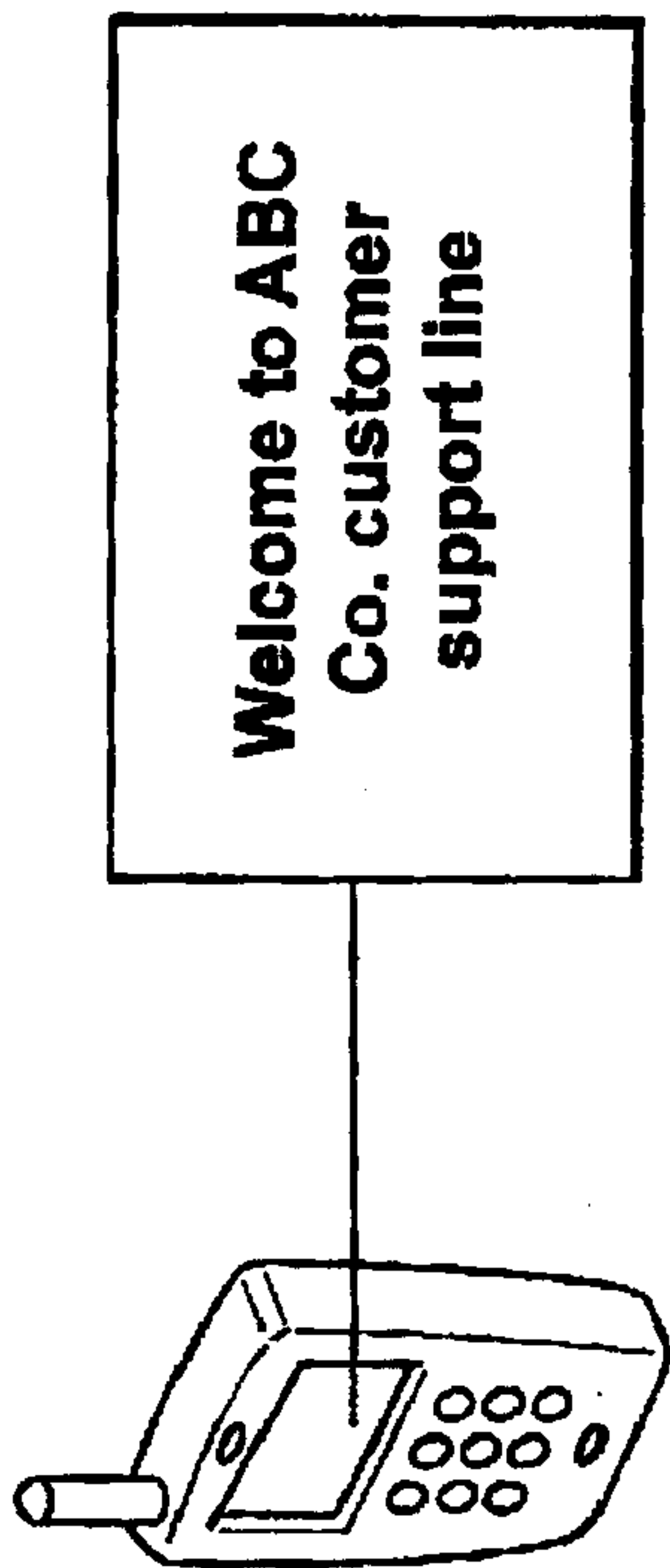


FIG. 1A

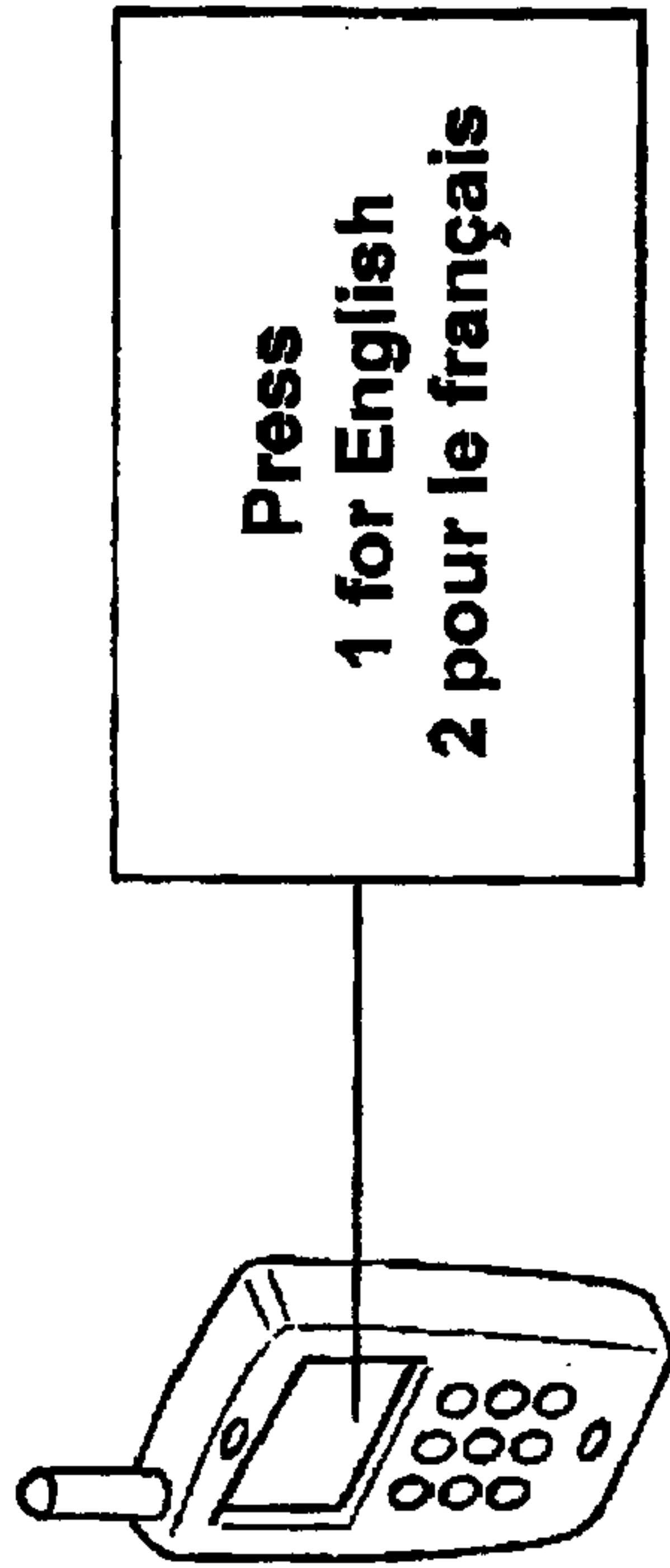


FIG. 1B

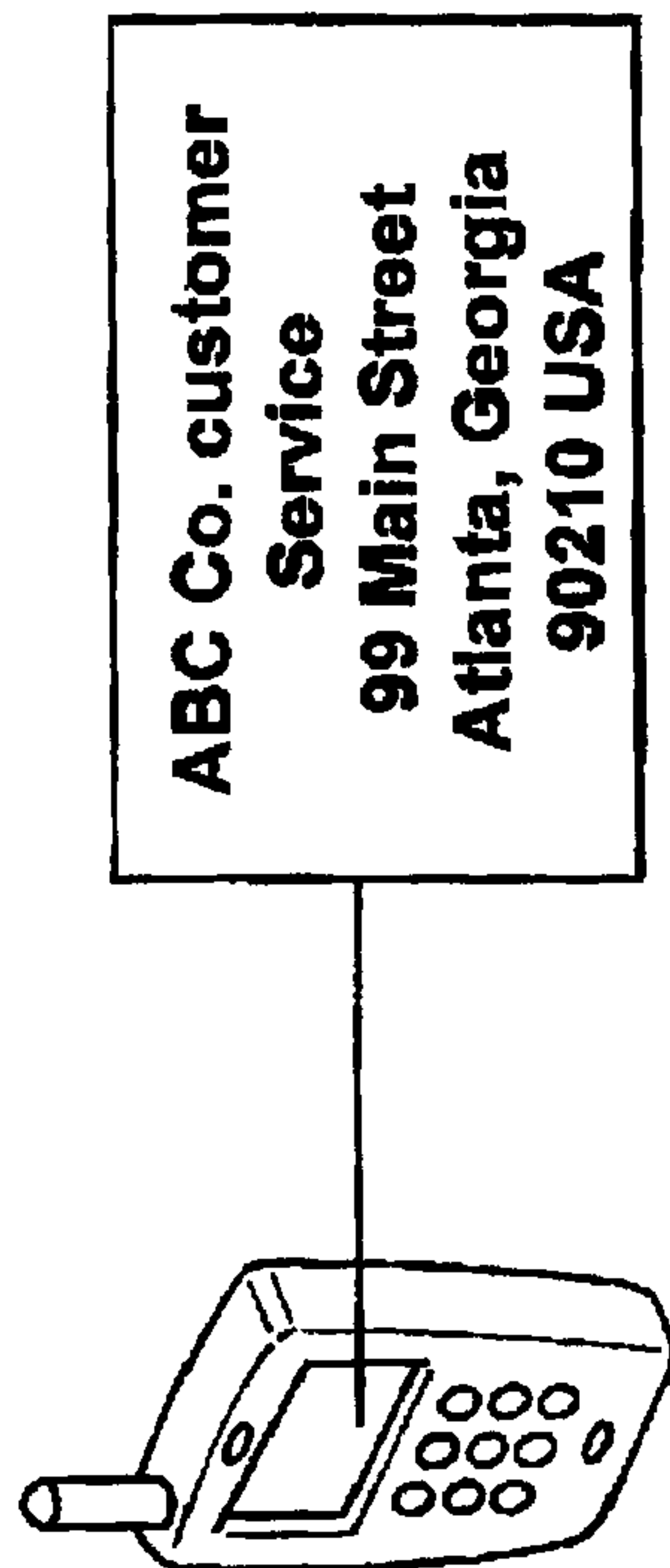


FIG. 1C

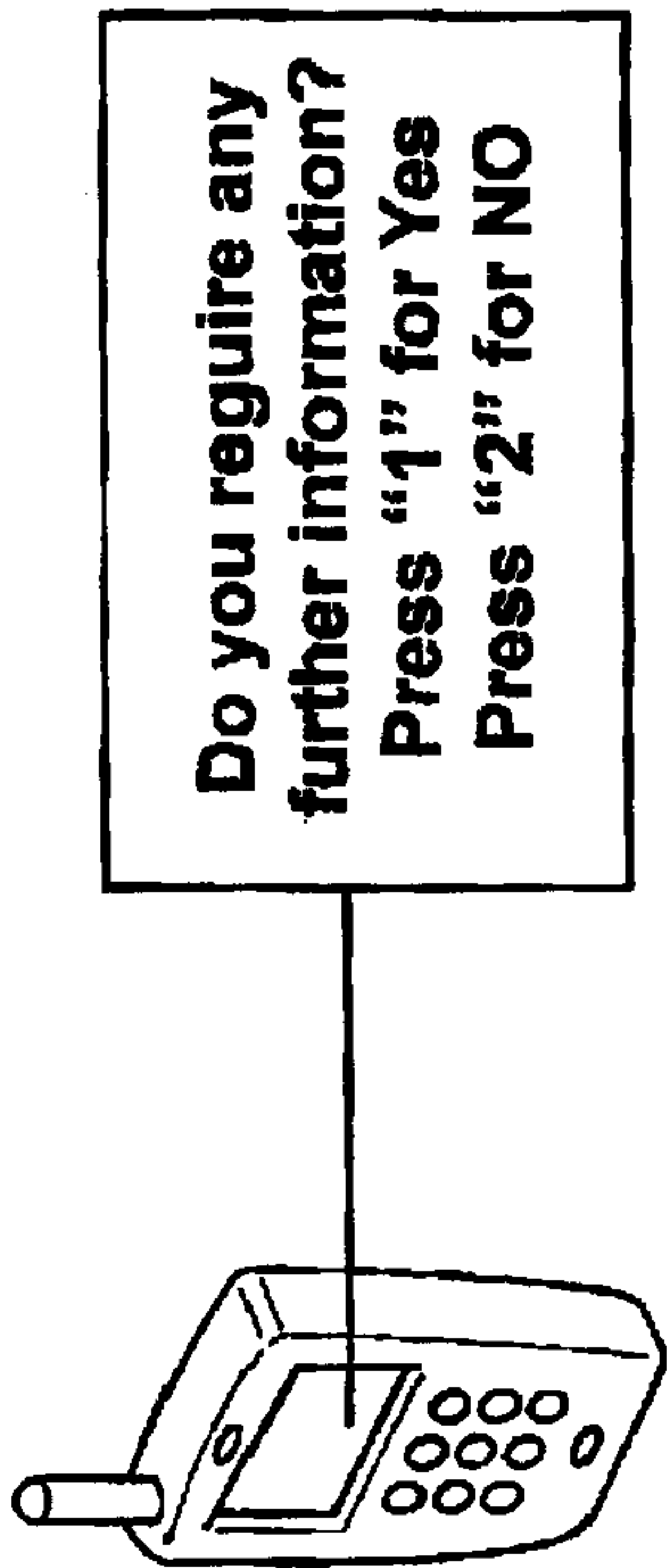


FIG. 1D

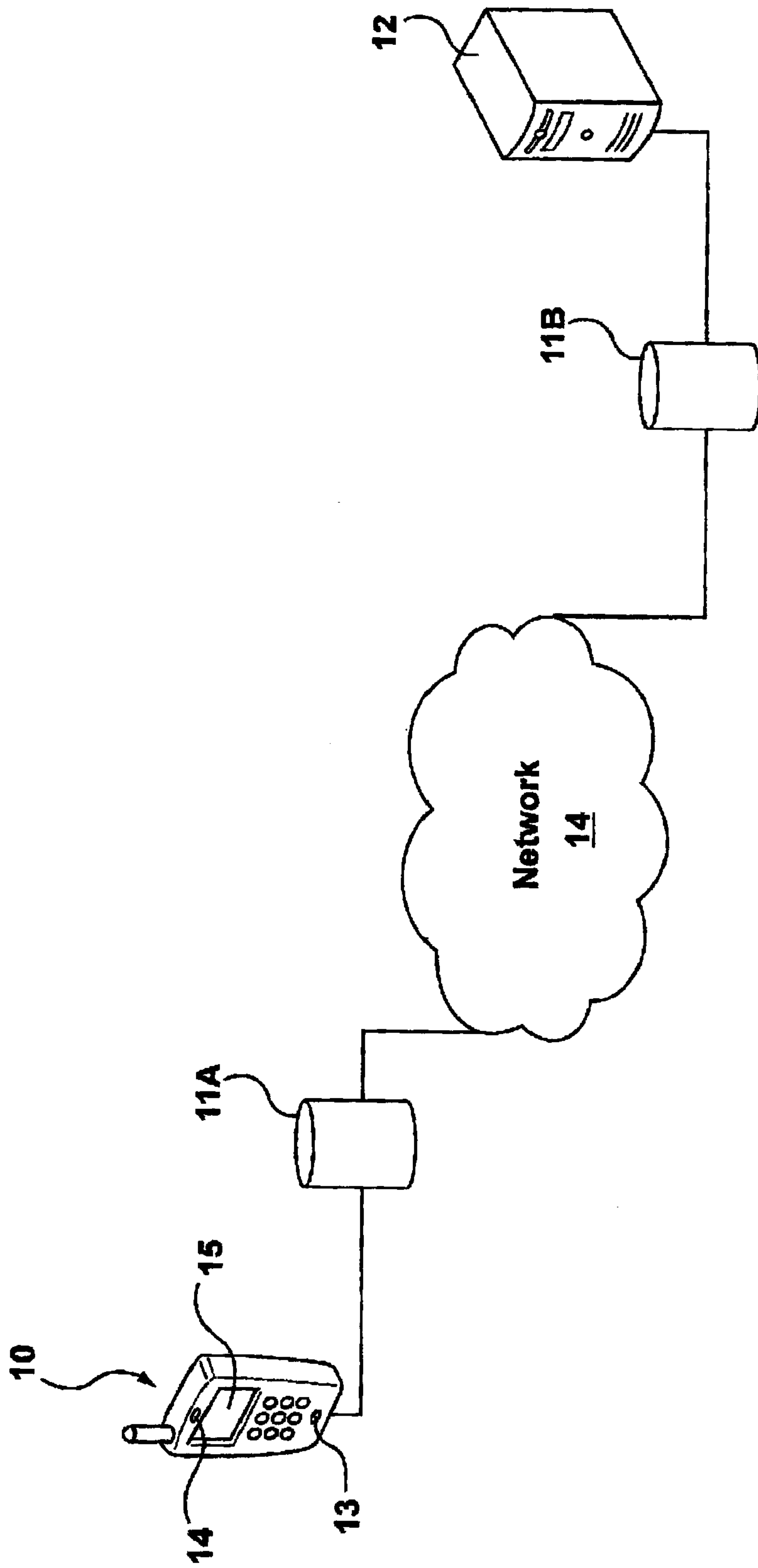


FIG. 2

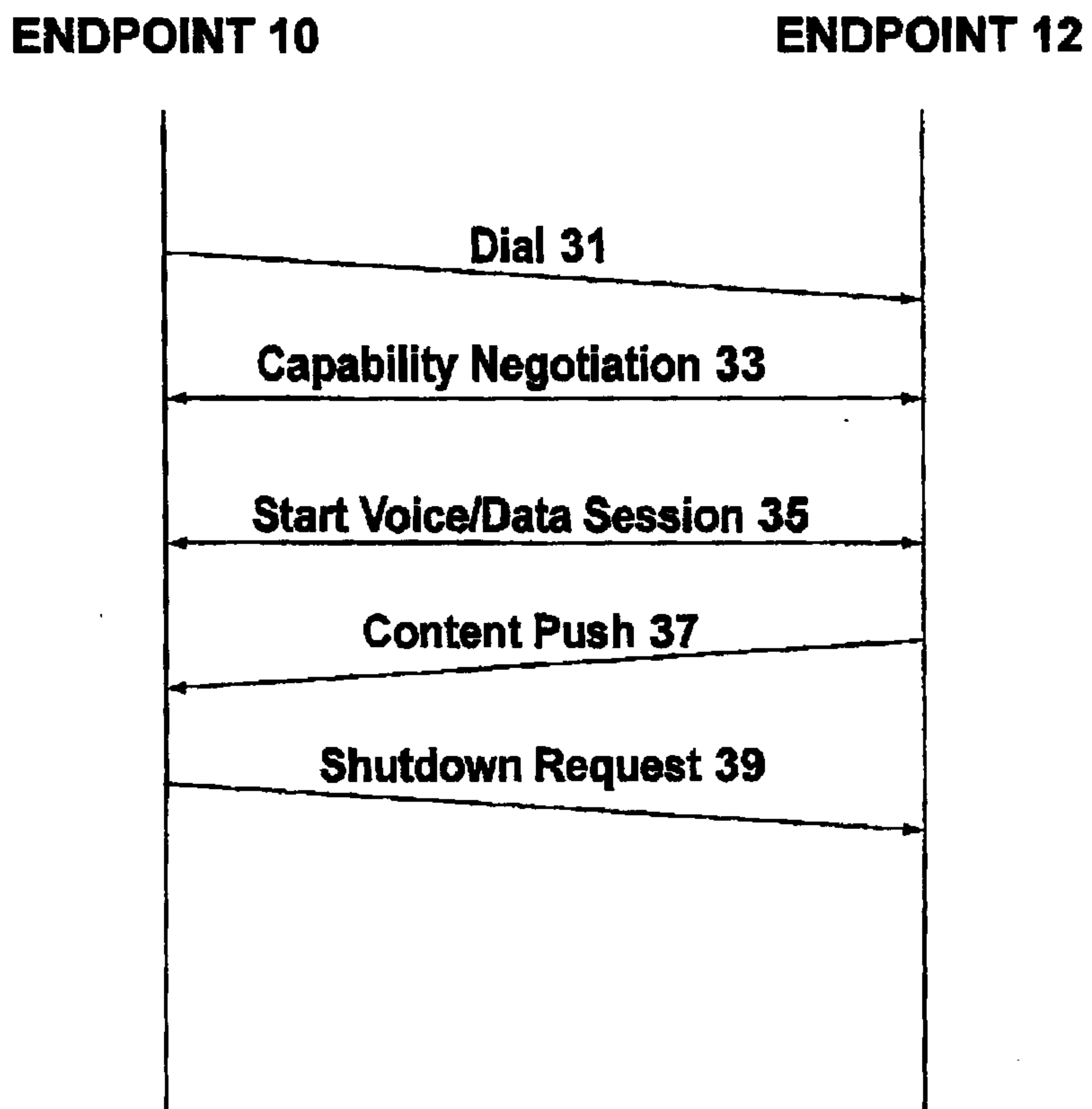


FIG. 3

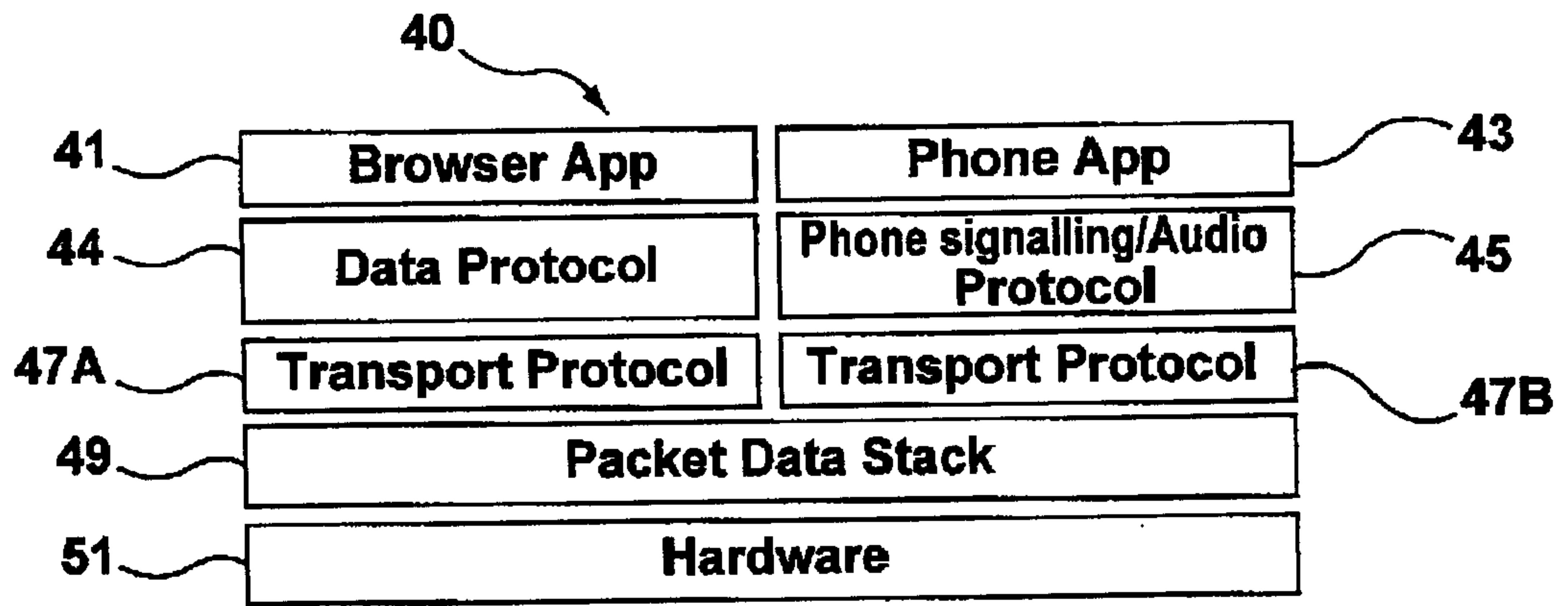


FIG. 4A

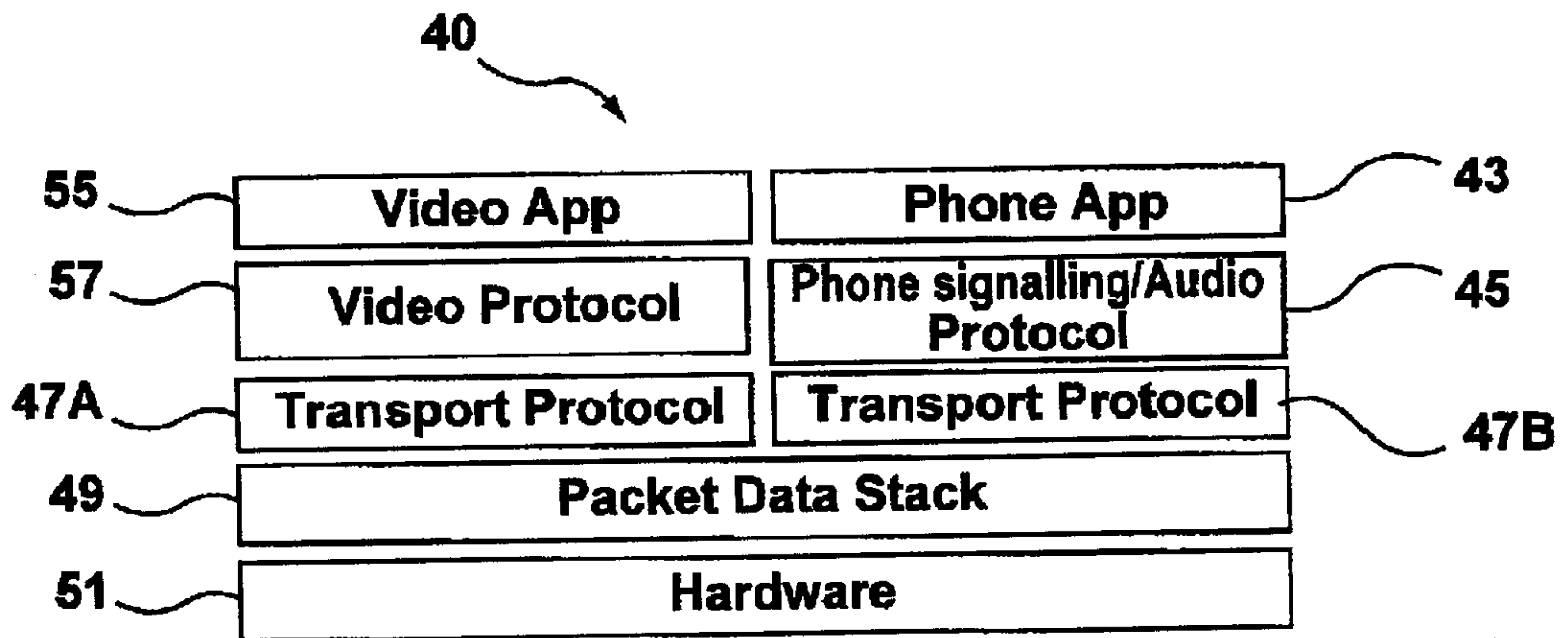


FIG. 4B

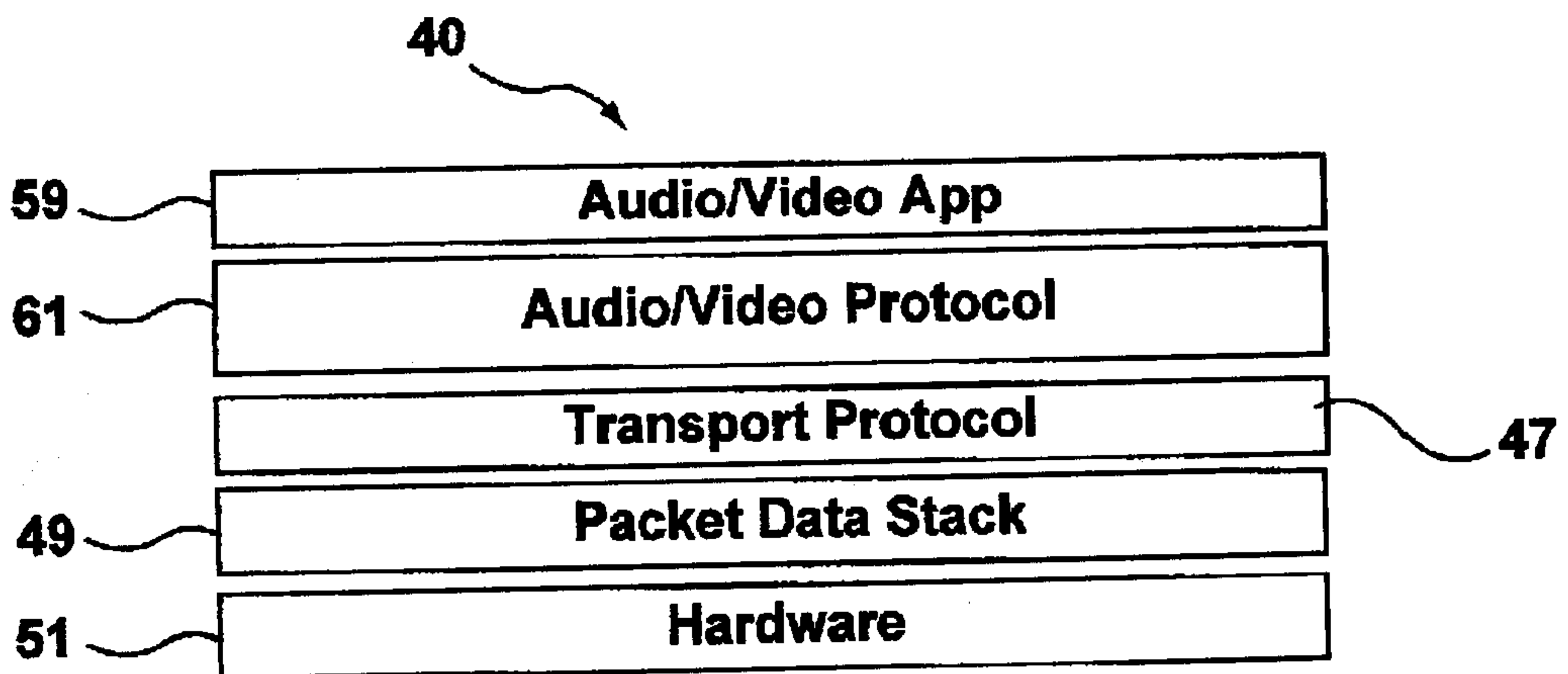


FIG. 4C

ENDPOINT 10

ENDPOINT 12

