METHOD AND APPARATUS FOR COUPLING A VISUAL BROWSER TO A VOICE BROWSER

BACKGROUND OF THE INVENTION

Technical Field

This invention relates to browsing network-based electronic content and more particularly to a method and apparatus for coupling a visual browser to a voice browser.

Description of the Related Art

Visual Browsers are applications which facilitate visual access to network-based electronic content provided in a computer communications network. One type of Visual Browser, the Web Browser, is useful for locating and displaying network-based electronic content formatted using HyperText Markup Language ("HTML"). Two popular Web Browsers are Netscape® Navigator® and Microsoft® Internet Explorer®.Notably, the term "Visual Browser" denotes that the browser can display graphics, text or a combination of graphics and text. In addition, most Visual Browsers can present multimedia information, including sound and video, although some Visual Browsers can require plug-ins in order to support particular multimedia information formats.

Whereas typical Visual Browsers operate in the desktop environment, compressed HTML ("C-HTML") Visual Browsers have emerged for processing HTML formatted documents in low-bandwidth environments. Specifically, C-HTML formatted documents are HTML formatted documents which have been compressed prior to transmission. C-HTML compliant Visual Browsers can decompress C-HTML formatted documents prior to displaying the same. Exemplary C-HTML Visual Browsers have been implemented for the QNX® Neutrino® operating system manufactured by QNX Software Systems, Ltd. of Kanata, Ontario.

A Voice Browser, unlike a Visual Browser, does not permit a user to interact with network-based electronic content visually. Rather, a Voice Browser, which can operate in conjunction with a Speech Recognition Engine and Speech Synthesis Engine, can permit the user to interact with network-based electronic content audibly. That is, the user can provide voice commands to navigate from network-based electronic document to document. Likewise, network-based electronic content can be presented to the user audibly, typically in the form of synthesized speech. Thus, Voice Browsers can provide

voice access and interactive voice response to network-based electronic content and applications, for instance by telephone, personal digital assistant, or desktop computer.

Significantly, Voice Browsers can be configured to interact with network-based electronic content encoded in VoiceXML. VoiceXML is a markup language for distributed voice applications based on extended markup language ("XML"), much as HTML is a markup language for distributed visual applications. VoiceXML is designed for creating audio dialogs that feature synthesized speech, digitized audio, recognition of spoken and Dual Tone Multifrequency ("DTMF") key input, recording of spoken input, telephony, and mixed-initiative conversations. Version 1.0 of the VoiceXML specification has been published by the VoiceXML Forum in the document Linda Boyer, Peter Danielsen, Jim Ferrans, Gerald Karam, David Ladd, Bruce Lucas and Kenneth Rehor, *Voice eXtensible Markup Language (VoiceXML™) version 1.0*, (W3C May 2000), incorporated herein by reference. Additionally, Version 1.0 of the VoiceXML specification has been submitted to and accepted by the World Wide Web Consortium by the VoiceXML Forum as a proposed industry standard.

Notably, the capabilities of Visual Browsers have not been combined with the capabilities of Voice Browsers such that a user of both can interact with network-based electronic content concurrently. That is, to date no solution has been provided which permits a user to interact with network-based visual content in a Visual Browser while also interacting with network-based audio content in a Voice Browser. Present efforts to provide a browser which can interact with network-based visual and audio content have been confined to the coding of speech synthesis functionality into an existing Visual Browser to produce a speech-aware Visual Browser. In addition, new speech-related markup tags for Visual Browsers have been proposed in order to provide speech functionality to a Visual Browser.

Still, these solutions require the implementor to develop a speech-aware function set for handling network-based speech content and to integrate the same directly in the source code of the Visual Browser. In consequence, the development of speech-related functionality is tightly linked to the development of the remaining functionality of the Visual

Browser. Finally, the tight integration between the Visual Browser and the speech-aware functionality precludes the user from using a separate, more robust and efficient Voice Browser having a set of functions useful for interacting with network-based speech content. Hence, what is needed is a method and apparatus for coupling a visual browser to a voice browser so that the combination of the visual browser and the voice browser can perform concurrent visual and voice browsing of network-based electronic content.

SUMMARY OF THE INVENTION

The present invention is a method and apparatus for concurrently accessing network-based electronic content in a Voice Browser and a Visual Browser. The method can include the steps of retrieving a network-based document formatted for display in the Visual Browser; identifying in the retrieved document a reference to the Voice Browser, the reference specifying electronic content formatted for audible presentation in the Voice Browser; and, transmitting the reference to the Voice Browser. The Voice Browser can retrieve the specified electronic content and audibly present the electronic content. Concurrently, the Visual Browser can visually present the network-based document formatted for visual presentation in the Visual Browser.

In a preferred embodiment of the present invention, the network-based document formatted for visual presentation in the Visual Browser can be an HTML document and the Visual Browser can be a Web Browser. Likewise, the electronic content formatted for audible presentation in the Voice Browser can be VoiceXML formatted electronic content. In the case where the network-based document is an HTML document, the reference can be a coordination markup attribute.

Significantly, the reference can specify a network-based document containing the electronic content formatted for audible presentation in the Voice Browser. Consequently, the Voice Browser can retrieve the specified network-based document through the computer communications network and audibly present the electronic content contained in the specified network-based document, while the Visual Browser can visually present the network-based document formatted for visual presentation.

In an alternative embodiment, a method for concurrently accessing network-based electronic content in a Voice Browser and a Visual Browser can include the steps of retrieving a network-based document formatted for audible presentation in the Voice Browser; identifying in the retrieved document a reference to the Visual Browser, the reference specifying electronic content formatted for visual presentation in the Visual Browser; and, transmitting the reference to the Visual Browser. The Visual Browser can retrieve the specified electronic content and visually present the specified electronic

content. Concurrently, the Voice Browser can audibly present the network-based document formatted for audible presentation in the Voice Browser.

More particularly, the network-based document formatted for audible presentation in the Voice Browser can be a VoiceXML document. Moreover, the electronic content formatted for visual presentation in the Visual Browser can be HTML formatted electronic content and the Visual Browser can be a Web Browser. Significantly, the reference can specify a network-based document containing the electronic content formatted for visual presentation in the Visual Browser. In that case, the Visual Browser can retrieve the specified network-based document through the computer communications network and visually present the electronic content contained in the network-based document, while the Voice Browser can audibly present the network-based document formatted for audible presentation.

A multi-modal browser for performing the method of the invention can include a Visual Browser for visually presenting visual content retrieved from a computer communications network, a Voice Browser for audibly presenting voice content retrieved from the computer communications network, and a Coupling Interface for synchronizing the visual and audible presentation of the visual and voice content in each of the respective Visual and Voice Browsers. Notably, the visual content can be HTML formatted content and the voice content can be VoiceXML formatted content.

Additionally, the multi-modal browser can include a Time Critical Event Handler. The Time Critical Event Handler can listen for time critical content, identify in the time critical content a content-type indicator, forward the time critical content to the Visual Browser if the content-type indicator indicates visual content, and forward the time critical content to the Voice Browser if the content-type indicator indicates voice content.

A method for modifying a network-based document for supporting concurrent access to network-based voice and visual content in a Voice Browser and a Visual Browser can include the steps of incorporating visual content in a network-based document; formatting the network-based document for visual presentation in a Visual Browser; and, inserting at least one markup tag in the network-based document. The

markup tag can contain a coordination markup attribute specifying a network-based document formatted for audible presentation in a Voice Browser. As a result, the Visual Browser when rendering the network-based document formatted for visual display can identify the coordination markup attribute in the markup tag and can transmit a reference to the specified network-based document to the Voice Browser. Consequently, the Voice Browser can retrieve the specified network-based document and audibly present the specified network-based document concurrently with the visual presentation of the network-based document formatted for visual presentation in the Visual Browser. In the preferred method, the network based document formatted for visual presentation in a Visual Browser can be an HTML formatted document and the Visual Browser can be a Web Browser. Also, the network based document formatted for audible presentation in a Voice Browser can be a VoiceXML formatted document and the Voice Browser can be a VoiceXML Browser.

A method for modifying a network-based document for supporting concurrent access to network-based voice and visual content in a Voice Browser and a Visual Browser can also include the steps of: incorporating voice content in a network-based document; formatting the network-based document for audible presentation in a Voice Browser; and, inserting at least one markup tag in the network-based document. The markup tag can contain a coordination markup attribute specifying a network-based document formatted for visual presentation in a Visual Browser. As such, the Voice Browser when rendering the network-based document formatted for audible display can identify the coordination markup attribute in the markup tag and can transmit a reference to the specified network-based document to the Visual Browser. In consequence, the Visual Browser can retrieve the specified network-based document and visually present the specified network-based document concurrently with the audible presentation of the network-based document formatted for audible presentation in the Voice Browser. In a preferred embodiment of the method, the network based document formatted for visual presentation in a Visual Browser can be an HTML formatted document and the Visual Browser can be a Web Browser. Also, the network based document formatted for

audible presentation in a Voice Browser can be a VoiceXML formatted document and the Voice Browser can be a VoiceXML Browser.

BRIEF DESCRIPTION OF THE DRAWINGS

There are presently shown in the drawings embodiments which are presently preferred, it being understood, however, that the invention is not limited to the precise arrangements and instrumentalities shown.

Figure 1 is a schematic representation of a server computer system supplying electronic content formatted for visual presentation in a Visual Browser in a client computer system, and a server computer system supplying electronic content formatted for audible presentation in a Voice Browser in the client computer system.

Figure 2 is a hierarchical illustration of an architecture suitable for use in the network of Figure 1 in which the client computer system can couple a Visual Browser and a Voice Browser so that both can concurrently access network-based electronic content in the network.

Figure 3 is a multi-modal browser architecture for implementing the method of the invention.

DETAILED DESCRIPTION OF THE INVENTION

The present invention is a method and apparatus for coupling a Visual Browser and a Voice Browser so that both can concurrently access network-based electronic content in a computer communications network in a coordinated and complementary fashion. The method and apparatus of the invention provide for a multi-modal aspect in consequence of which the functions of visual and voice browsing can be synchronized without developing a new, separate hybrid browser and without having to code new speech handling methods directly into existing visual browsers. Rather, an independent Visual Browser can be coupled to an independent Voice Browser through a reference embedded in a network-based document retrievable by the Visual Browser. Conversely, an independent Voice Browser can be coupled to an independent Visual Browser through a reference embedded in a network-based document retrievable by the Voice Browser. In both cases, existing tags defined by existing markup languages are utilized rather than the modification of browser functionality in order to facilitate the coupling of the Voice Browser and the Visual Browser.

In the preferred embodiment, a network-based document formatted for display in a Visual Browser can be retrieved from a source in the computer communications network. A reference to a Voice Browser can be identified in the retrieved document in which the reference specifies electronic content formatted for audible presentation in the Voice Browser. Subsequently, the reference can be transmitted to the Voice Browser. Responsive to receiving the transmitted reference, the Voice Browser can extract therefrom the specified electronic content and audibly present the electronic content while the Visual Browser concurrently visually presents the electronic content contained in the network-based document.

Figure 1 illustrates a network environment in which the method of the invention can be performed. This network environment comprises a computer communications network 10 interconnecting a client computers 1 and servers 20, 30 where the servers 20, 30 include at least one voice content server and at least one visual content server. Significantly, the invention is not limited in regard to the number of server computers

required to provide the client computer 1 with network-based voice content and network-based visual content. In fact, in one embodiment of the present invention, a single server computer can provide both voice and visual content to the client computer 1. Still, a server computer separately providing voice and visual electronic content to requesting client computers are shown for ease of illustration and discussion.

The computer communications network 10 can be any non-publicly accessible network such as a LAN (local area network) or WAN (wide area network), or preferably, the Internet. In operation, the client computer 1 can establish a connection with server computers 20, 30 in order to transmit a request for electronic content stored in the server computers 20, 30. Each server computer 20, 30 can reply to the request by supplying the requested electronic content to the client computer 1. Notably, the electronic content can be formatted for visual presentation in a Visual Browser or audible presentation in a Voice Browser.

In the preferred embodiment, the electronic content formatted for visual presentation in a Visual Browser is an HTML formatted electronic content which can be displayed in a Web Browser. Similarly, in the preferred embodiment, the electronic content formatted for audible presentation in a Voice Browser is VoiceXML formatted electronic content. Still, the invention is not limited to the particular method of formatting the electronic content. Rather, any suitable method for formatting visual electronic content and audible electronic content can suffice. Examples of suitable alternative visual formatting methods can include C-HTML, XML, SGML, VRML, etc. Similarly, examples of suitable alternative audible presentation methods can include WAV, MPEG, AIFF, etc.

As shown in Figure 1, the client computer 1 preferably has a central processing unit (CPU) 2, an internal memory device 3 such as a random access memory (RAM), and a fixed storage 4 such as a hard disk drive (HDD). The client computer 1 also includes network interface circuitry (NIC) 5 for communicatively connecting the client computer 1 to the computer communications network 10. Optionally, the client computer 1 can further include a keyboard 7 and at least one user interface display unit 6 such as a video display terminal (VDT) operatively connected thereto for the purpose of interacting with the client

computer 1. For the purposes of rendering the voice content, the client computer 1 need not even have the keyboard 7, a mouse or display unit 6 as they are unnecessary for interacting with a user through voice. Additionally, the client computer 1 that interacts audibly with the user can contain audio circuitry, a speaker and a microphone (not shown). Further, the client computer 1 need not be a personal computer as illustrated in Figure 1. Rather, the client computer 1 can be any computing device containing such audio circuitry and enabled to communicate with the computer communications network 10 in order to perform the methods of the invention.

Like the client computer 1, server computers 20, 30 preferably have CPUs 22, 32, internal memory devices 23, 33, and fixed storage 24, 34. Each server computer 20, 30 also includes network interface circuitry (NIC) 25, 35 for communicatively connecting the server computers 20, 30 to the computer communications network 10. Unlike the client computer 1, the server computers 20, 30 need not include a keyboard or video display terminal inasmuch as the server computers can interact with users over a network. In the preferred embodiment, server computer 20 can provide to the client computer 1 visual content for display in a Visual Browser residing in the client computer 1. Concurrently, server computer 30 can provide to the client computer 1 voice content for audible presentation in a Voice Browser in the client computer 1. Still, the invention is not limited in the precise number of server computers supplied to provide electronic content to the client computer 1. Rather, in an alternative embodiment, both the visual content and the voice content can be provided to the client computer 1 by a single server computer.

Turning now to Figure 2, a hierarchical illustration of an architecture suitable for use in the network of Figure 1 is shown in which the client computer 1 can couple a Visual Browser 12 and a Voice Browser 13 so that both can concurrently access network-based electronic content in servers 20, 30 residing in the computer communications network 10. As shown in Figure 2, each of the server computers 20, 30 can include an operating system 21, 31 upon which applications can execute. As is well-known in the art, the operating system 21, 31 can be used to provide network communications to an application executing thereon.

The server computer 20 can include a server application for responding to requests for visually presentable electronic content. More particularly, the server application can provide requested electronic content in response to received requests. In the preferred embodiment, the server application is a Web Server 26 which can provide HTML formatted content to requesting computers communicatively connected to the computer communications network 10. Specifically, the HTML formatted content can be static HTML Web pages stored in the server computer 20, or dynamically formatted HTML Web pages created dynamically in response to received requests. Alternatively, the HTML formatted content can be a compressed HTML document.

Similarly, the server computer 30 can include a server application for responding to requests for audibly presentable electronic content. More particularly, like the Web Server 26 of server computer 20, the server application of server computer 30 can provide requested electronic content in response to received requests. In the preferred embodiment, the server application is a VoiceXML Server 36 which can provide VoiceXML formatted content to requesting computers communicatively connected to the computer communications network 10. Specifically, the VoiceXML formatted content can be static VoiceXML documents stored in the server computer 30, or dynamically formatted VoiceXML documents created dynamically in response to received requests.

In the preferred embodiment, the client computer 1 can include an operating system 11, for example QNX Neutrino, upon which applications can load and execute. Two such preferred applications can include a Visual Browser 12 and a Voice Browser 13. The Visual Browser 12, for instance a Web Browser, can accept user requests for Web content in the Web server 26 communicatively connected to the computer communications network 10. In response, the Visual Browser 12 can transmit to the Web server 26 the request using the HyperText Transfer Protocol ("HTTP"). The Web server 26 can receive the request and can transmit the requested Web content to the Visual Browser 12. Similarly, the Voice Browser 13 can accept user requests for VoiceXML content in the VoiceXML Server 36 communicatively connected to the computer communications network 10. Notably, the user requests can be speech recognized by

the speech recognition engine 15. In response, the Voice Browser 13 can transmit to the VoiceXML Server 36 the speech recognized request also using HTTP. The Voice Server 36 can receive the request and can transmit the requested VoiceXML content to the Voice Browser 13. Subsequently, the audibly presentable electronic content in the VoiceXML page can be audibly presented to the user by way of speech synthesis engine 14. In this way, a user of client computer 1 can both view visual content using a Visual Browser 12 and hear voice content using a Voice Browser 13.

Significantly, a user of client computer 1 can "co-browse" both visual and voice content provided by the Web Server 26 and VoiceXML Server 36, respectively. That is, the visual content browsing of the Visual Browser 12 can be synchronized with the voice content browsing of the Voice Browser 13 by coupling the visual browsing functionality of the Visual Browser 12 to the voice browsing functionality of the Voice Browser 13. This coupling can be accomplished without requiring the design and implementation of a hybrid content browser. Rather, the Visual Browser 12 can be coupled to the Voice Browser 13 by means of coordination markup attributes contained in existing markup tags included as part of the HTML and VoiceXML specification.

In particular, in the preferred embodiment HTML formatted content and VoiceXML formatted content can include "co-target" type tags designed to reference other browsers. Specifically, in an HTML formatted document, an anchor reference markup tag containing an "href" can be used to specify target Web content to be loaded when a user selects the anchor reference. Typically, the target Web content is specified by a uniform resource locator ("URL") specified in the href. In the preferred embodiment, a co-target specifying VoiceXML content, known as a "cohref", can be embedded in the anchor reference in addition to the target Web content. An example of a co-target specified by a cohref follows:

```
<HTML>
    <HEAD>
        <TITLE>SBU CoBrowsing Demo</TITLE>
    </HEAD>
    <BODY LEFTMARGIN="0" TOPMARGIN="0">
```

```
            <A HREF="mainmenu.html" cohref="mainmenu.vxml">
            <IMG SRC="splashscreen.jpg" WIDTH="314" HEIGHT="200"/></A>
        </BODY>
    </HTML>
```

As shown in the above example, the hyperlink specified by the anchor markup includes a reference both to a target Web page specified by "mainmenu.html" and a co-target VoiceXML document specified by "mainmenu.vxml". When the Visual Browser 12 encounters this hyperlink, the Visual Browser 12 can request from the Web Server 26 the Web page "mainmenu.html". Additionally, the Visual Browser 12 can identify the co-target reference and can transmit the reference to the Voice Browser 13. The Visual Browser 12 can transmit the reference to the Voice Browser 13 using the visit() method contained in the published applications programming interface ("API") of the Voice Browser 13. In response, the Voice Browser 13 can request from the VoiceXML Server 36 the VoiceXML document "mainmenu.vxml".

Correspondingly, in a VoiceXML formatted document, a "choice next" markup tag can be used to specify target VoiceXML content to be loaded when a user vocally selects a particular choice in the VoiceXML document. Typically, the target VoiceXML content is specified by a uniform resource locator ("URL") specified in the choice next tag. In the preferred embodiment, a "conext" attribute specifying Web content can be embedded in the choice next tag in addition to the target VoiceXML content. An example of a choice next tag incorporating a conext attribute follows:

```
<vxml version="1.0">
    <form name="Welcome">
        <block>Welcome to the CoBrowsing Demo!
            <goto next="#1"/>
        </block>
    </form>
    <menu id="1">
        <prompt>Say test to authenticate user.</prompt>
        <grammar type="text/jsgf">
            continue
```

```
            </grammar>
            <choice next="mainmenu.vxml" conext="mainmenu.html">
                test
            </choice>
        </menu>
    </vxml>
```

As shown in the above example, the hyperlink specified by the next tag includes a reference both to a target VoiceXML document specified by "mainmenu.vxml" and a co-target Web page specified by "mainmenu.html". When the Voice Browser 13 encounters this next, the Voice Browser 13 can request from the Voice Server 36 the VoiceXML document "mainmenu.vxml". Additionally, the Voice Browser 13 can identify the conext reference and can transmit the reference to the Visual Browser 12. The Voice Browser 13 can transmit the reference to the Voice Browser 12 using a published API function call of the Visual Browser 12 for requesting a Web page by the URL of the Web page, for example the HTTP "Open" function. The Visual Browser 12, in turn, can request from the Web Server 26 the Web page "mainmenu.html".

Significantly, the present invention is not limited merely to the href and conext coordination markup attributes. Rather, a number of coordination markup attributes are available for use in the Visual Browser 12 and the Voice Browser 13 for coordinating the actions of the Visual Browser 12 with the actions of the Voice Browser 13. More particularly, the attributes can be divided into two categories. First, attributes can reflect actions which occur at the source of an event, such as choosing a voice menu item, completing a voice form, or clicking on a visual link or button. Second, attributes can reflect actions which occur at the destination of an event, such as visiting a voice body, menu or form, or displaying a visual document or anchor within a document. Notably, with regard to the Visual Browser 13, coordination markup attributes can be defined which are not a standard part of the HTML specification. In consequence, those non-standard attributes can be defined in a customizable visual browser having a corresponding augmented tag attribute set. In the preferred embodiment, the following table presents preferred coordination markup attributes for use in the Visual Browser 12 and the Voice

Browser 13 of the present invention:

	At Source	At destination
Speech	<goto next="URL" conext="URL">	<menu covisit="URL">
Browser	<choice next="URL" conext="URL">	<form covisit="URL">
	<form action="URL" coaction="URL">	
Visual	<a href="URL" cohref="URL">	<body covisit="URL">
Browser	<form action="URL" coaction="URL">	<form covisit="URL">
		<a id="id" covisit="URL">

With regard to the Speech Browser 13, "conext" specifies the URL for the Visual Browser 12 to display when an associated voice-specified goto or menu choice is selected by a user. "Coaction" specifies the URL for the Visual Browser 12 to display when an associated voice form is completed. Finally, "covisit" specifies the URL for the Visual Browser 12 to visit when an associated voice element is visited. Similarly, with regard to the Visual Browser 12, "cohref" specifies the URL for the Voice Browser 13 to visit when an associated visual link or button is selected. "Coaction" specifies the URL for the Voice Browser 13 to visit when an associated visual form is submitted. Finally, "covisit" specifies the URL for the Voice Browser 13 to visit when an associated visual document is loaded or a visual anchor is displayed.

Notably, both the visual and voice content can be included in a single, composite network-based document. In the case of a composite network-based document, both the Voice Browser 13 and the Visual Browser 12 can parse the composite document and present content contained therein which is suitable for display in the respective browser. Additionally, each of the Voice Browser 13 and the Visual Browser 12 can identify in the composite document, respective cohref and conext specified electronic content. As example of a composite document follows:

```
<HTML>
    <HEAD><TITLE>SBU Embedded Markup CoBrowsing Demo</TITLE></HEAD>

    <vxml version="1.0">
```

```
                <form name="Welcome">
                    <block>Welcome to the SBU Multi-modal Browser Demo!
                        <goto next="#1"/>
                    </block>
                </form>
                <menu id="1">
                    <prompt>Say test to authenticate user.</prompt>
                    <grammar type="text/jsgf">
                        continue
                    </grammar>
                    <choice next="mainmenu.vxml" conext="mainmenu.html">
                        test
                    </choice>
                </menu>
            </vxml>

            <BODY LEFTMARGIN="0" TOPMARGIN="0">
                <A HREF="mainmenu.html" cohref="mainmenu.vxml">
                    <IMG SRC="splashscreen.jpg" WIDTH="314" HEIGHT="200"/>
                </A>
            </BODY>
        </HTML>
```

Figure 3 illustrates a multi-modal browser 150 for implementing the method of the invention. As will be apparent from the figure, a multi-modal browser 150 in accordance with the inventive arrangements can include three distinct components: a Visual Browser 120, a Voice Browser 130, and a Coupling Interface 100 for coupling the Voice Browser 130 to the Visual Browser 120 by managing the synchronization between each browser. Additionally, the multi-modal browser 150 can include a Time Critical Event Handler 140 for processing time critical information such as status information. Significantly, the Visual Browser 120 and Voice Browser 130 components are independent. Each interprets its own markup stream to present independent, but complementary information to a user 160. In accordance with the method of the invention, the browsers can interact with one

another through the Coupling Interface 100.

In the preferred embodiment of the multi-modal browser 150, the Visual Browser 120 is a Java-based Micro Browser designed for execution in an embedded environment. The Micro Browser can be implemented using VisualAge® for Embedded Systems® manufactured by IBM Corporation of Armonk, New York. VisualAge for Embedded Systems supplements the Java user interface support with Degas® display technology. In particular, Degas is a Java-based development environment used for building GUI embedded applications for execution in the QNX Neutrino embedded operating system. Degas can be used to create screen displays having unique graphical views and a realistic 3D look-and-feel. Specifically, a developer utilizing the Degas framework can construct the visual elements for a target device, subclass the classes that implement those elements, and prototype a GUI interface using those elements.

The Micro Browser can extend the Degas DefaultApplication class and implements an HTML View/Controller pair as an extension of the Degas DefaultView and DefaultController classes, respectively. The HTML Controller can pass user events from the display to the Micro Browser application, which in turn can update the HTML View. A more detailed discussion of the Degas framework can be found in Kerri Ann Hamberg and Brian Neal, *GUI Navigation in Embedded Automotive Systems*, http://www.tradespeak.com/htmldocs/2030.html (tradespeak 2000), incorporated herein by reference.

As shown in Figure 3, another multi-modal browser 150 component is the Time Critical Event Handler 140. The Time Critical Event Handler 140 can listen on an IP port for markup which can be forwarded to either the Visual Browser 120, the Voice Browser 130, or both via the Coupling Interface 100. The markup corresponds to time critical information (referred to as status information) which may interrupt the current application. In order to determine which browser should display the status information, the markup preferably includes a content-type, or MIME type, placed at the head of the document containing the markup. According to convention, two blank lines follow the content-type above either the beginning HTML or VXML document tag. In the preferred embodiment,

markup having status information which is received in the multi-modal browser 150 preferably has a content-type of VXML, CHTML, or EVXML.

In operation, based on a user's interaction with voice or visual content, the user 160 can navigate to a new location. As each browser transitions from one URL to another, the browser can coordinate browsing with the other browser via coordination markup attributes added to the respective markup languages. These element attributes can direct the browser to notify the other browser when to visit a complementary URL ("co-URL"). For example, when the user 160 selects a link in the Visual Browser 120, the Visual Browser 120 visits the URL of the visual page specified by the link. To direct the Voice Browser 130 to visit the corresponding voice page, an application developer can add a covisit attribute containing the URL of the desired voice page to the HTML anchor or form tag in the visual page. Similarly, when the user 160 selects a link in the Voice Browser 130, the Voice Browser visits the URL of the voice page specified by the link. To direct the Visual Browser 120 to visit the corresponding visual page, the application developer can add a comenu, coform, or conext attribute containing the URL of the desired visual page to the VXML tag in the voice page.

Thus each browsing component (voice and visual) can interpret its own markup stream and make its own transition from one document or document element (URL) to another document or document element. The document elements contain markup attributes that synchronize the transitions of each browser to the other, at points chosen by the application designer.

Thus, coordination markup attributes, for example cohref and conext tags, can be used to couple a Visual Browser to a Voice Browser without requiring the wholesale modification of either. Rather, the content markup, itself, can act as the coupling agent. Hence, as an example, when a user mouse-clicks on a visual display of a map in a Visual Browser, a co-target tag associated with the map can indicate to a Voice Browser to load and playback audio content associated with the map. Conversely, a Voice Browser can prompt a user to speak the name of a geographic location for which the Voice Browser can audibly provide further information. In response to receiving speech input specifying a

geographic location, a co-target tag associated with the voice prompt can indicate to a Visual Browser to load and display a map showing the location of the specified geographic location.

CLAIMS

1. A method for concurrently accessing network-based electronic content in a Voice Browser and a Visual Browser comprising the steps of:

retrieving a network-based document formatted for display in the Visual Browser;

identifying in the retrieved document a reference to the Voice Browser, said reference specifying electronic content formatted for audible presentation in the Voice Browser; and,

transmitting said reference to the Voice Browser;

the Voice Browser retrieving said specified electronic content and audibly presenting said electronic content in the Voice Browser;

the Visual Browser visually presenting said network-based document concurrently with said audible presentation.

2. The method according to claim 1, wherein said network-based document formatted for visual presentation in the Visual Browser is an HTML formatted document and the Visual Browser is a Web Browser.

3. The method according to claim 1, wherein said electronic content formatted for audible presentation in the Voice Browser is VoiceXML formatted electronic content.

4. The method according to claim 2, wherein said reference is a coordination markup attribute.

5. The method according to claim 1, wherein said reference specifies a network-based document containing said electronic content formatted for audible presentation in the Voice Browser,

whereby the Voice Browser can retrieve said specified network-based document through the computer communications network and audibly present said electronic content contained in said specified network-based document, while the Visual Browser

can visually present said network-based document formatted for visual presentation.

6. A method for concurrently accessing network-based electronic content in a Voice Browser and a Visual Browser comprising:

retrieving a network-based document formatted for audible presentation in the Voice Browser;

identifying in the retrieved document a reference to the Visual Browser, said reference specifying electronic content formatted for visual presentation in the Visual Browser; and,

transmitting said reference to the Visual Browser;

the Visual Browser retrieving said specified electronic content and visually presenting said electronic content in the Visual Browser;

the Voice Browser audibly presenting said network-based document concurrently with said visual presentation.

7. The method according to claim 6, wherein said network-based document formatted for audible presentation in the Voice Browser is a VoiceXML document.

8. The method according to claim 6, wherein said electronic content formatted for visual presentation in the Visual Browser is HTML formatted electronic content and the Visual Browser is a Web Browser.

9. The method according to claim 6, wherein said reference specifies a network-based document containing said electronic content formatted for visual presentation in the Visual Browser,

whereby the Visual Browser can retrieve said specified network-based document through the computer communications network and visually present said electronic content contained in said network-based document, while the Voice Browser can audibly present said network-based document formatted for audible presentation.

10. A machine readable storage, having stored thereon a computer program for concurrently accessing network-based electronic content in a Visual Browser and a Voice Browser, said computer program having a plurality of code sections executable by a machine for causing the machine to perform the steps of:

retrieving a network-based document formatted for display in the Visual Browser;

identifying in the retrieved document a reference to the Voice Browser, said reference specifying electronic content formatted for audible presentation in the Voice Browser; and,

transmitting said reference to the Voice Browser;

the Voice Browser retrieving said specified electronic content and audibly presenting said electronic content in the Voice Browser;

the Visual Browser visually presenting said network-based document concurrently with said audible presentation.

11. The machine readable storage according to claim 10, wherein said network-based document formatted for visual presentation in the Visual Browser is an HTML document and the Visual Browser is a Web Browser.

12. The machine readable storage according to claim 10, wherein said electronic content formatted for audible presentation in the Voice Browser is VoiceXML formatted electronic content.

13. The machine readable storage according to claim 11, wherein said reference is a coordination markup attribute.

14. The machine readable storage according to claim 10, wherein said reference specifies a network-based document containing said electronic content formatted for audible presentation in the Voice Browser,

whereby the Voice Browser can retrieve said specified network-based document through the computer communications network and audibly present said electronic content contained in said specified network-based document, while the Visual Browser can visually present said network-based document formatted for visual presentation.

15. A machine readable storage, having stored thereon a computer program for concurrently accessing network-based electronic content in a Visual Browser and a Voice Browser, said computer program having a plurality of code sections executable by a machine for causing the machine to perform the steps of:

retrieving a network-based document formatted for audible presentation in the Voice Browser;

identifying in the retrieved document a reference to the Visual Browser, said reference specifying electronic content formatted for visual presentation in the Visual Browser; and,

transmitting said reference to the Visual Browser;

the Visual Browser retrieving said specified electronic content and visually presenting said electronic content in the Visual Browser;

the Voice Browser audibly presenting said network-based document concurrently with said visual presentation.

16. The machine readable storage according to claim 15, wherein said network-based document formatted for audible presentation in the Voice Browser is a VoiceXML document.

17. The machine readable storage according to claim 15, wherein said electronic content formatted for visual presentation in the Visual Browser is HTML formatted electronic content and the Visual Browser is a Web Browser.

18. The machine readable storage according to claim 15, wherein said reference

specifies a network-based document containing said electronic content formatted for visual presentation in the Visual Browser,

whereby the Visual Browser can retrieve said specified network-based document through the computer communications network and visually present said electronic content contained in said specified network-based document, while the Voice Browser can audibly present said network-based document formatted for audible presentation.

19. A multi-modal browser comprising:

a Visual Browser for visually presenting visual content retrieved from a computer communications network;

a Voice Browser for audibly presenting voice content retrieved from said computer communications network; and,

a Coupling Interface for synchronizing said visual and audible presentation of said visual and voice content in each of said respective Visual and Voice Browsers.

20. The multi-modal browser of claim 19, wherein said visual content is HTML formatted content and said voice content is VoiceXML formatted content.

21. The multi-modal browser of claim 19, further comprising:

a Time Critical Event Handler,

said Time Critical Event Handler listening for time critical content,

said Time Critical Event Handler identifying in said time critical content a content-type indicator,

said Time Critical Event Handler forwarding said time critical content to said Visual Browser if said content-type indicator indicates visual content,

said Time Critical Event Handler forwarding said time critical content to said Voice Browser if said content-type indicator indicates voice content.

22. A method for modifying a network-based document for supporting concurrent

access to network-based voice and visual content in a Voice Browser and a Visual Browser, comprising the steps of:

incorporating visual content in a network-based document;

formatting said network-based document for visual presentation in a Visual Browser; and,

inserting at least one markup tag in said network-based document, said markup tag containing a coordination markup attribute specifying a network-based document formatted for audible presentation in a Voice Browser,

whereby the Visual Browser when rendering said network-based document, formatted for visual display can identify said coordination markup attribute in said markup tag and can transmit a reference to said specified network-based document to the Voice Browser causing the Voice Browser to retrieve said specified network-based document and audibly present said specified network-based document concurrently with the visual presentation of said network-based document formatted for visual presentation in the Visual Browser.

23. The method according to claim 22, wherein said network based document formatted for visual presentation in a Visual Browser is an HTML formatted document and said Visual Browser is a Web Browser.

24. The method according to claim 22, wherein said network based document formatted for audible presentation in a Voice Browser is a VoiceXML formatted document and said Voice Browser is a VoiceXML Browser.

25. A method for modifying a network-based document for supporting concurrent access to network-based voice and visual content in a Voice Browser and a Visual Browser, comprising the steps of:

incorporating voice content in a network-based document;

formatting said network-based document for audible presentation in a Voice

Browser;

inserting at least one markup tag in said network-based document;

said markup tag containing a coordination markup attribute specifying a network-based document formatted for visual presentation in a Visual Browser,

whereby the Voice Browser when rendering said network-based document formatted for audible display can identify said coordination markup attribute in said markup tag and can transmit a reference to said specified network-based document to the Visual Browser causing the Visual Browser to retrieve said specified network-based document and visually present said specified network-based document concurrently with the audible presentation of said network-based document formatted for audible presentation in the Voice Browser.

26. The method according to claim 25, wherein said network based document formatted for visual presentation in a Visual Browser is an HTML formatted document and said Visual Browser is a Web Browser.

27. The method according to claim 25, wherein said network based document formatted for audible presentation in a Voice Browser is a VoiceXML formatted document and said Voice Browser is a VoiceXML Browser.