SURFING THE VOICE WEB: ISSUES IN THE DESIGN OF A
VOICE BROWSER

Michael H. Cohen
Vice President, Dialog R&D
Nuance Communications, 1380 Willow Road, Menlo Park, CA 94025
mcohen@nuance.com

ABSTRACT

Wireless access to the internet has emerged as a potential "killer application" which will demand sophisticated spoken language technology in order to realize its full potential. There are many challenges that must be met to make the "Voice Web" a reality – recognition technology, network infrastructure, new devices, and the development of new user interface paradigms. The Voice Web offers an opportunity to create a unified user interface to the entire world of services and communications. This talk will present the challenges presented by the Voice Web from the user interface perspective, and describe Voyager, a voice browser developed by Nuance Communications.

1. INTRODUCTION

Wireless access to the internet has emerged as a huge opportunity, driving a great deal of technology development and infrastructure investment. One of the biggest limitations, thus far, when accessing the web from small, mobile devices has been the user interface – keyboards are impractical for small devices, and pointing and clicking has limited expressive power. Speech technology is emerging as a core piece of the user interface for wireless web access. Many believe this is the long awaited “killer app” for speech technology.

The actual opportunity is much bigger than just web access. Adding speech to the interface provides the opportunity to unify the telephone network and the internet – creating a new paradigm uniting the kind of communications we are accustomed to over the telephone network with the information and transaction access we expect from the World Wide Web.

Nuance has created a voice browser, called Voyager, in order to provide a consistent interface for users to the entire world of voice-aware applications and services as well as the communications capabilities of the telephone network. The biggest challenges in creating the browser have been at the level of the voice user interface – how can we provide an easy-to-use, easy-to-learn, consistent interface to the diverse world of independently developed applications and services? Section 2 will describe some of the voice user interface (VUI) design challenges in building a browser for the Voice Web. Section 3 will describe a number of features of Voyager designed to meet these challenges.

2. VUI DESIGN CHALLENGES

There are a number of characteristics of voice interfaces, in general, that pose challenges to the design of effective, usable systems. The nature of voice browsing introduces some new challenges, and amplifies some of the general difficulties.

One of the biggest issues, for all VUIs, is controlling the cognitive load on the caller. Purely auditory interfaces challenge human memory and attention due to the strictly serial nature of information presentation to the user. Unlike screen-based interfaces, which can create constant reminders for users (e.g., with toolbars or icons on the screen), sound is evanescent. There are no constant reminders that can be referenced whenever needed – therefore, a level of control is taken away from the user. As a result, auditory interfaces challenge human memory and learning in ways that screen-based interfaces do not.

Voice browsing adds additional cognitive load for a number of reasons. First of all, context tracking is far more complex than in a single application. At any stage of the dialog, the caller must know the context they are in, and the available commands and functionality. They may be talking to the browser itself or to any of the applications or services that can be accessed through the browser (henceforth, the set of applications and services available through the browser will be referred to as “content sites.”)
Each content site may be designed by a different developer, have different capabilities, or even use different vocabulary to refer to similar things. Context can change quickly (e.g., if hyperlinking from one content site to another).

Furthermore, the number of commands and concepts that must be learned is far greater than for the typical application. Callers must be able to control many different applications, as well as the browser. This complexity can be partly mitigated by well designed closed voice portals that create a cohesive set of services that inter-operate well, but the problem is much harder to solve for open web browsing.

Another challenge for VUI design is efficiency. Surveys of users of deployed spoken language systems suggest that callers care about efficiency – they are unhappy when completing seemingly simple operations takes a long time, or many steps. The need for efficiency is amplified when browsing – in which case many transactions may be performed in a single phone call. This will be especially true when voice browsers become the “intelligent dialtone” of the future. Users will not tolerate perceived inefficiencies whenever they pick up the telephone.

Finally, care must be taken to make sure that the user experience does not become tedious and repetitious with frequent use. Many browser-based services are being deployed with a subscription model – subscribers are expected to call frequently and use the system for extended periods. Some techniques that may be “tolerable” for a one-time caller quickly become annoying or frustrating with repeated use. For example, a main menu prompt which is repeated frequently, in exactly the same way, may become annoying. As new business models emerge in which service providers wish to keep callers on the line longer to hear more ads, perform more transactions, or burn more cellular minutes, techniques must be developed to keep the caller interested and engaged.

With these VUI design challenges in mind, the following design goals were chosen for Voyager:

- **Minimize Cognitive Load**: Minimize the short term memory load on the caller, the complexity of concepts the caller must understand, and the number of things they must learn in order to successfully use the system.

- **Maximize Efficiency**: Facilitate rapid transaction completion and quick navigation to content sites of interest.

- **Maximize User Control**: Provide the facilities to the caller to control and optimize their browsing experience, step in and out of content, request information and help whenever needed, and maintain constant access to Voyager capabilities.

Create a Sustainable User Experience: Develop a system that “wears well over time” by creating an effective system persona that is engaging and drives the crafting of conversational prompt wording and variation.

### 3. VOYAGER FEATURES

Voyager includes many features which have become part of standard “best practice” VUI design, such as the availability of universal commands callers can use to stay on track or resolve confusion (e.g., “help”, “repeat”), extensive context-specific help facilities, error recovery mechanisms, etc. These features help reduce cognitive load, increase efficiency, and provide the user with greater control of the progress of the call and a means to help themselves if they run into problems. This section will focus on Voyager features particular to enhancing the browsing experience, while also helping to achieve our design goals and maximize overall system usability.

Most of the features of Voyager can be divided into two major categories: navigation and personalization. The navigational features allow the caller to easily locate content sites and to visit content sites and place calls. Navigational features include hotword, voicelinks, traversal commands, and bookmarks.

- **Hotword**: The hotword is a word or phrase (e.g., “Voyager”) the caller can say at any time (even if in the middle of some content or on a phone call) to get the attention of the browser. Once the hotword is recognized, the entire universal grammar of the browser is available to the caller. The ability to use hotword to get the browser’s attention gives the caller enhanced control over the session – they have access to all of the browser’s capabilities at all times.

- **Voicelinks**: Voicelinks are the analog to hyperlinks on the World Wide Web. Voicelinks can be set up to transfer a caller to a different content site, or elsewhere within the same content site.

- **Traversal commands**: The browser tracks the sites a caller has visited within a session. The caller can use the commands “go back” or “go forward” to revisit sites.
**Bookmarks:** The caller can save a bookmark to any site. In the future, the caller can visit the site just by saying the bookmark name to the browser. This makes use of the system far more efficient.

The personalization features include explicit methods by which the caller can customize the browser to their needs (e.g., by creating a personal profile) and methods by which the browser automatically adapts to the caller’s needs (e.g., just-in-time instruction). Personalization features include profiles, verification, just-in-time instruction, and bookmarks.

**Profiles:** The caller can enter personal information that can be used by content sites to complete transactions. This greatly increases efficiency, and saves the caller from repetitious and tedious tasks.

**Verification:** Callers register their voiceprint with the system. This can be used by the browser to assure the identity of a caller before sharing profile information.

**Just-in-Time Instruction:** One of the biggest challenges in the design of the browser is to limit the amount of material callers must learn in order to use the system. This is a challenge, in general, with systems that offer lots of functionality. Some systems use tutorials to describe to a first-time caller all of the capabilities of the system. If the tutorial covers more than a small amount of material, there is little chance the caller will recall the material when they need it. The principle behind just-in-time instruction is to teach a very small amount of material at any one time, and to teach a caller something when they need it, so that they immediately exercise the new knowledge. The system tracks the caller’s history, and plays just-in-time instructions the first time a caller tries to use a feature, if the caller is running into problems, or if the caller is using a feature inefficiently.

**Bookmarks:** Bookmarks are mentioned again here because, in addition to being a navigational aid, they are a means by which the caller can personalize the browser to their needs.

### 3. CONCLUSIONS

Voyager is a browser of the Voice Web which presents the caller with a consistent interface to voice sites and to the telephone network. The talk will demonstrate a number of features of Voyager, and show how they help achieve the design goals described above.

---

**ACKNOWLEDGMENTS**

Jennifer Balogh, Trace Wax, James Giangola, Melissa Dougherty, and Nicole Leduc all played key roles in the planning, design, and testing of the Voyager user interface.