VoiceXML
A Summary of Recent Writing on VoiceXML

The article, " VoiceXML: The New Standard in Web Telephony" is a basic discussion on implementing a VoiceXML solution for telephony systems. VoiceXML is a product for communicating voice data, as both value and behavior, between applications. As one of the distant progeny of the now aging Extensible Mark-Up Language (XML), Voice XML provides a standardized mark up that follows the lexical lead of SGML (standardized generalize markup language). The VoiceXML model is partially supported by the World Wide Web Consortium (Voice Browser Activity).

This article is primarily an instructional guide for creating a VoiceXML implementation. It also provides information about DTMF (Dual Tone Multi Frequency) keypad standards for telephone enabled web access and the benefits of speech recognition. The DTMF keypad standards are highlighted to provided VoiceXML authors with a background of DTMF conventions so that they may include them in their design. Voice communications are described as superlative to DTMF menu selection, but the author admits that DTMF menu selection is a good fallback for users that are uncomfortable with or unable to respond through verbal communication. Phillips is careful to mention "how much simpler it would be to allow users to request information using natural language." The article also reminds us that, "speech recognition fulfills that need"(Phillips). The discussion of DTMF keypad standards is limited to navigational and confirmational user command. The keypad standards described include: "'1' for 'yes' and '2' for 'no'," "the star key . . . to exit a menu," the "operator key . . . to connect the caller to a human operator, " and " the pound key . . . to start over at the beginning of a menu" (Phillips). These "generally accepted standards" (Phillips) are offered by the author as guidelines and admittedly contain "quite a lot of flexibility" (Phillips).

The majority of the article focuses on an imaginary application that provides the context from which design and coding are discussed. This application allows a user to access "email, a local weather report, or the thought for the day " (Phillips) over the telephone. The design scope for this application is limited to "the logic that allows you [the designer] to choose among these options" (Phillips).

The remaining sections of the article communicate the specific syntax of the VoiceXML product. These sections introduce the basics of "coding the application" (Phillips). Our author exposes the syntax of grammar for "external grammar" files and "explicit inline grammars" (Phillips). The option and prompt tags are also discussed.

In general, the article is myopic and prescribes a quick solution for implementation. The discussion of design is disappointingly inadequate for even the most fundamental exposition of this technology. It fails to describe important issues in design, modeling, and implementation.

The author frequently mentions "this Version of VoiceXML " but never explicitly defines what version it is. In explaining the option tag, the author states, "an option list can easily handle DTMF " but that "the version we are using doesn't support this construct either" (Phillips). It is unclear if the version described is a

current release, an old release, or discusses respective of the authors expectations of future releases.

The author's myopia is most clearly represented in their limited discussion of the technological philosophy behind VoiceXML. Most developers would notice the fundamental oddity in the creation of Voice XML. The benefits of XML are in data transport and its straightforward, non-specific syntax. Adopting a more specific voice XML standard goes against the principles of good XML design. To better adhere to the standard set forth by the W3C it would behoove a developer to create an XML document that is transformed via XSLT into the required markup. This allows the developer the flexibility to change their presentation layers as the technology changes and offers the desired separation of application, presentation, and data layers. The author admits, "Few existing implementations are fully compatible . . . because the standard has been either subsetted or extended" (Phillips). If the standard does not exist or is unstable, it makes even more sense to abstract this variable layer from the core design. Instead of designing VoiceXML documents, developers should be creating XML documents that transform into any needed voice technology's data format. That is not to say that an attempted at a standard mark up language for speech should not be developed. It seems more beneficial to set the focus of a VoiceXML model toward a standard mark up language, not an extensible one.

Ironically, this article treats Voice XML simply as a voice markup language, not an extensible voice mark up language, as the acronym reads. Without the extensibility provided by its ties to XML this language would prove only slightly more useful than Cascading Style Sheets with Aural Capabilities (part of the CSS-2 W3C recommendation) for client side application, or a well defined standard mark-up for voice applications. What the article fails to champion is the very benefit of creating a Voice XML standard. The ability to create an open architecture affords the advanced benefits the author defines as missing from computer audio recognition. Any XML related technology, should by design, afford developers the ability to make new data types. If for example, the Voice XML Schema used by an organization needs to detect sarcasm it can. Discussion of extensibility is too clearly missing from this article. The W3C VoiceXML recommendation does describe a language that is extensible by nature, but the author does not mention this fact.

An additional XML related shortfall of the article is its inadequate discussion of validation schemas. There is discussion of using DTD (Document Type Definition), but there is no discussion of writing references for less antiquated validation schemas like XDR (XML Data Reduced). This is peculiar in discussing a burgeoning technology, since it seems logical that new technologies should be mated with new technologies. If the VoiceXML product only supports DTD schemas then it seems like a good idea to mention this limitation.

Discussion of the benefits of this technology to the disabled is also missing from the article. This same publication offered an article in the January issue that expounded the need for such development on web sites and developed tools("Web Accessibility" Jan. 2001). The W3C prologue to their recommendations on Voice communications reads that the organization " is working to expand access to the Web". They explicitly state that this technology "will be a boon to people with visual

impairments" (W3C). It seems as though such benefits are of equal if not greater importance than being able" to buy a Lear Jet" or "detect sarcasm" (Phillips) over the telephone.

The author also fails to site the benefits of marketing access to an entirely new class of customers. That new market would be people unable or unwilling to purchase computers. This is an additional benefit that the W3C sites in their preliminary discussion of voice technologies.

Despite the author's indirection, there are some clear lessons to be learned from the walkthrough. It does introduce the syntax of a VoiceXML document. The author's step-by-step example is straightforward, if not complete. It may be speculated that the author refrained from widening the scope of the article to meet publishing or reader guidelines. Its lack of key information for even an introductory article may be attributed to the instability of the technology's standards at the time the article was published. Perhaps the author should revisit the technology now that a W3C document has been published about it.

Works Cited

Paciello, Michael G.. "Web Accessibility: 500 Million and Growing."

Webreview (Jan 2001):

< http://www.webreview.com/2001/01_16/developers/index01.shtml >

Phillips, Lee Anne. "VoiceXML: The New Standard in Web Telephony."

Webreview (Feb 2001):

< http://www.webreview.com/2001/02_09/developers/index01.shtml >

"Voice Browser Activity — Voice enabling the Web!" World Wide Web Consortium Oct 2001

< http://www.w3.org/Voice/ >