

- SHOPPING CART
- LOG IN
- MEDIA CENTER
- MAILING LIST
- CONTACT US
- HOME

[EVENTS](#)
[TRAINING](#)
[CONSULTING](#)
[NEWSLETTERS](#)
[WEBCASTS](#)
[BLOGS](#)

VoiceXML 2.0 -- Sharper Across the Board

Free CallCenter Insider Newsletter

Your Email Address

Optional Areas of Interest

- International New s
- Advice/Tips
- Technology
- Agent Development
- IVR

- [Magazine](#)
- [New Customer Management Insight](#)
- [Digital Publication](#)
- [Call Center Magazine](#)
- [Archives](#)
- [FREE Webcasts](#)
- [Podcasts](#)
- [Vendor Guide](#)
- [Multimedia Content](#)
- [Online Advice](#)
- [eNewsletter Alerts](#)
- [ICMI Forum](#)
- [ICMI Jobs](#)
- [Training](#)
- [Consulting](#)
- [Newsletters](#)
- [Publications/Tools](#)
- [Industry Research](#)
- [QUEUETIPS](#)
- [Conferences](#)
- [ICMI Global Report](#)
- [Call Center Knowledge Online](#)
- [Join ICMI](#)
- [Members Only](#)
- [Vendor Guide](#)
- [Industry Information](#)
- [About ICMI](#)
- [Download the ICMI Catalog](#)

VoiceXML 1.0 laid the foundation for leveraging web resources in support of voice applications. Now, Version 2.0 extends the developer's control: providing sharper, more incisive tools for vocabulary- and grammar-building, and for fine-tuning the behavior of the voice interface, itself.

By Andy Green and Ellen Muraskin

02/06/2002, 10:26 AM ET

The VoiceXML markup language, a subject seen on these pages for about three years now, is supposed to do for IVR what HTML did for the visual web: promote a community of developers who will "publish" information to phone callers as easily as visual webmasters publish to web servers. Push the very same information, in fact: order status, stock quotes, flight arrivals, and the like.

Like HTML pages, VoiceXML "pages" can be linked to and retrieved via HTTP protocol from servers that may be located far from the actual browser, creating a truly roamable "voice web" instead of an IVR platform that dead-ends a phone number. Unlike HTML, the "browser" in this case is not on the end-user side: It's the VoiceXML interpreting software, and the set of navigation and personalization commands, plus the speaker-rec and text-to-speech algorithms that "listen" and "speak" to callers. Since the caller to a VoiceXML page only has speech and touch-tone for interactivity, he has on the CPE side what's equivalent to the keyboard, mouse, and screen of Netscape or IE, not the browser programs themselves.

Like HTML, VoiceXML has spawned a raft of web server platforms and browsers, and gone through subsequent releases. Version 1.0 of VoiceXML was submitted to the W3C (World Wide Web Consortium) in May 2000 by the specification's original proponents and the nucleus of the Voice-XML Forum, consisting of AT&T, Lucent, Motorola, and IBM. Version 2.0, the fruit of the W3C's Voice Browser Working group (which now includes some 60 companies), was released late this past October and is now open for review. A finalized "recommendation" is expected by late 2002.

WRITE ONCE, HOST ANYWHERE

► UTILITIES

- [print this article](#)
- [e-mail this article](#)

► RELATED LINKS

[Roundtable: The Quest For Voice QoS](#)

[Sponsor Link: The Strategic and Financial Justifications for Convergence](#)

► TRAINING

► [View seminar calendar](#)

TechEncyclopedia

Define It

As with HTML, a more fully developed VoiceXML specification serves to make the language more functional and more portable, by reducing developers' need to code their own versions of such things as barge-in or dynamic grammars. Version 2.0 is furthest along in its standardization of text-to-speech controls and speech-recognition grammar formatting. It adds new rigor to the overall structure of describing dialogs. It is furthest behind in its specification of call control elements.

W3C Speech Interface Framework



Release 2.0 of the W3C's VoiceXML specification introduces subsidiary XML schemas for speech recognition grammars, speech synthesis controls, and other components of speech-enabled IVR.

Andrew Hunt is SpeechWorks' (Boston, MA - 617-428-4444, www.speechworks.com) Director of Standards and Emerging Technology, and a co-editor of the spec. The day after 2.0 was released for comment, he and Scott McGlashan, Leader of the 2.0 spec effort, and CTO of Swedish VoiceXML platform maker PipeBeach (Stockholm, Sweden - +46 708 46 24 32, www.pipebeach.com) were at SpeechTEK 2001 in New York, flush with accomplishment, presenting the standard and its new release.

At that session, Hunt noted that the VoiceXML forum went after several main areas in revising the standard. The first was "hailing platform and document interoperability." This involved resolving ambiguities in the spec, establishing minimum platform requirements in a VoiceXML interpreter platform, standardizing grammars and prompting, and establishing a testing protocol to prove conformance. The second thrust concerned feature enhancements, adding to what Voice-XML can simply do without resorting to lower-level code. A third goal was keeping it human-readable.

VoiceXML 2.0 now sits atop new subsidiary standards that deal with input (speech recognition) and output (text-to-speech). On the input side, The W3C Speech Recognition Grammar Format, based on Sun's Java speech rec grammar format, is acceptable to 2.0 in one of two forms: XML and Augmented BNF, a preexisting format that is more human-readable. Speech Synthesis Markup Language SSML (recommendation due Q3) is actually another XML application, as VoiceXML is, and is

based on Sun's Java speech Markup Language. Both the grammar XML and Speech Synthesis ML will have their own schemas/document type definitions (DTDs).

SPEECH SYNTHESIS MARKUP

SSML has replaced old proprietary prosody controls with new elements for voice-rendering paragraphs, sentences, normalizing dates and currencies, emphasis, choice of voice, pitch, break, and volume. It has set minimum recorded audio file support at 8 kHz, 8-bit mono PCM single-channel in a-law or m-law, raw and WAV formats; this last is the basis for concatenating template sentences and TTS into meaningful automated responses. Support for HTTP1.1 transport is now required for cache management and document management. For more specifics, see sidebar, [VoiceXML 2.0 Feature Enhancements](#). A more fully standardized VoiceXML will allow applications to be more easily ported from one TTS and ASR product to the other, and by the same token, from one host platform to the other. And it should also encourage the "loan" of common, application-specific grammars (think city names, think city names in Italy for Italian apps), perhaps retrieved through real-time links.

As of late this past October, only SpeechWorks, via their Speechify Text-to-Speech product, and a freeware TTS from Carnegie Mellon University supported SSML. And while all VoiceXML platform makers pledge compliance with 2.0, they can only adopt the parts that are near approval now; others will shake out this year. Some platform designers say they're waiting for a recommended 2.0 to come out before they release their first VoiceXML product.

DOUBLE-STANDARD BROWSERS

VoiceXML 1.0 and VoiceXML 2.0 pages will not work in the same application, nor on the same interpreter. McGlashan told his attendees, though, that a combination of automatic transformation and manual editing would convert documents without excessive expense. Further, one browser deployment is likely to support both interpreter versions, he said.

A subgroup of the W3C Voice Browser Working Group is now working on call control. According to McGlashan, they're now examining the telephony controls - call initiation, third-party conferencing - currently in the spec with the aim of making them protocol independent.

More complex call control capability, such as would manage multiple call legs in a rerouted application, may be put into a separate call control language. "We would then invite VoiceXML interpreters to participate in those calls," says McGlashan. He could give no timeline here, "since it's basically trying to reconstruct a lot of what telephony standards have been doing. It's a complex issue." The separate call control markup language already has a name, ccXML (call control XML). In the meantime, some middleware vendors have extended VoiceXML with proprietary call handling tags (Voice Genie) or else created a call handling markup language of their own (Voxeo's CallXML).

VOICEXML INTERPRETER PLATFORMS

The roundup that follows is representative. Others will follow by 2003, as the standard solidifies.

[1](#) | [2](#) | [3](#) | [4](#) | [Next Page](#) > >

									
									