

# CCXML: THE POWER OF STANDARDIZATION

**Authors:** Claudia Romellini, Flavio Tonelli (Loquendo)

**Date:** September 27, 2005.

## INTRODUCTION

This paper provides an introduction to CCXML (Call Control eXtensible Markup Language), a standard designed to power speech application deployment with advanced call control features whilst removing legacy from proprietary call control implementation. The article outlines the main features and benefits of CCXML and presents VoxNauta 7.0, the latest release of Loquendo's Speech Platform, which includes the implementation of a CCXML interpreter, compliant with the standard's 1.0. specification.

## CCXML BENEFITS & ADVANTAGES

Call Control extensible Mark-up Language (CCXML) is a language purposely designed for the advanced call control of phone calls. It is a W3C (World Wide Web Consortium) specification defined by the Voice Browser Working Group; the latest draft has been released in June 2005 [1] as a Last Call Working Draft.

CCXML was conceived by the World Wide Web Consortium to complement VoiceXML – in order to overcome certain limitations in VoiceXML call control related functionalities and process model, which were not appropriate for managing typically asynchronous telephony events.

Loquendo has actively contributed to the CCXML specification process and acting on its standard-adhering commitment has been one of the first speech companies to offer a CCXML outfitted vocal platform. (see the VoxNauta platform offer below).

CCXML offers an abstract layer for call control, independent from the underlying telephony system and completely “unaware” of the protocols used for the signalling and transport of voice calls. CCXML is the key factor for new cost-effective integrated voice services platforms easing the convergence process from traditional telephony (TDM) to voice-over-IP (VoIP), as it provides the same level of service to both networks.

Independence from telephony is accompanied by independence from the developing environment, as XML does not require specific tools and compilers, it is widely acknowledged and can be run on any platform. CCXML applications are fast and easy to develop, portable onto any platform and can be run by any CCXML compliant interpreter.

Combining the advantages of standardization and of an XML web-based approach, CCXML is becoming a *de facto* requirement because it allows portability and reusability, maintainability and cross-vendor interoperability, eliminating any awkward call control customization associated to specific vendors.

Although it was originally conceived for use in the speech systems area to complement VoiceXML, CCXML is now considered to fit the needs of any call-processing related application. For example, a PBX with ACD functionalities can be implemented based on this standard language, from a core of a few scripts to a complete framework of functionalities, ranging from priority-served queues to sophisticated conference support. Such a high level programmable

switch is well suited for dynamic configuration and per-call customisation, which until now could only be obtained through vendor modifications.

CCXML also brings many advantages for the development process, thanks to the split in the competencies required to implement dialogue and call management. The proper skills can now be optimally allocated, i.e. speech designers can focus on dialogue implementation and be relieved of the burden of intricate call control management, which can in turn be handled by a dedicated telephony competence team. Splitting the process into two domains will speed up all the phases of producing a speech application, from development to testing and maintenance.

CCXML favours efficiency at run time too; as the VoiceXML interpreter will only be committed to running the dialogues, the heavy burden of handling the telephonic events will be alleviated.

Both enterprise investments and customer satisfaction will benefit from the convergence of protocols and services in the new distributed, web-based and Vo-IP scenarios and CCXML is the means of accessing an infinite array of innovative open standards and open architecture solutions which we will shortly see in the telephony industry, an industry which until now has been very slow-changing.

## **NEW CCXML FEATURES: MAKING COMPLEX TASKS EASY**

The CCXML which addresses the specific scope of call control, empowers the functionalities available to application developers, and makes them accessible in a standardized uniform way, avoiding any vendor-specific customisation used in previous VoiceXML implementation:

- Handling asynchronous telephony events.
- Ability to associate a dedicated VoiceXML dialogue to each call leg.
- Concurrent execution of CCXML sessions and the ability to move the connections around from one session to another.
- Ability to send and receive events to and from external processes outside CCXML and VoiceXML platforms.
- Call set-up information handling for call screening and filtering, allowing intelligent routing and selective answering.
- Multiple-call handling capabilities: multi party conferencing, supervised transfer, whispering, call hold and outbound call.

A CCXML document is a collection of “handlers” which correspond to the states of the call. Leveraging a reduced number of powerful elements, this programming model facilitates the design of call handling applications. Developers have simplified access to all functionalities and can combine them using VoiceXML dialogues to build complex services and exploit a variety of new solutions which were difficult to imagine before.

## CCXML & VOICEXML FOR INNOVATIVE FEATURED APPLICATIONS

Two use cases will be taken into account in this article in order to illustrate the potential of a combined CCXML-VoiceXML phone and dialogue system, with some variants that differ from the classic IVR functionalities but can easily be implemented with CCXML, to substitute traditionally hard-to-configure systems.

### Enhanced Contact Center Automation

Let's consider an incoming call to a contact center that provides services for a number of companies. The CCXML root document recognizes the DNIS to be associated to an air company, then queries the *back-end* company's CRM systems with the CLI of the call to retrieve the caller data and the caller is found to be a registered user. With this information, the call is connected to a VoiceXML dialogue which will greet the user and ask him to choose from a menu of available services: flight schedules and timetable, booking, online check-in, special offers, frequent flyer miles & points.

Having obtained the desired information, the user asks to be put through to a *live* agent. As all the agents are busy, the user is connected to another dialogue rather than being simply put on hold with waiting music. The CCXML application inserts the call reference into a priority queue, as the user in question is entitled to receive *gold programme* priority service .

The dialogue, which is constructed according to the caller data, tells the user he's been promoted to the gold card programme thanks to the latest points earned and notifies him of his current mileage. While he waits, the user can interact with the dialogue application to hear the latest news or the company's promotions. He can also ask to be automatically called back as soon as an agent is ready to deal with a new call.

When an agent becomes available, the call related data is passed onto the agent's desktop, the connection to the dialogue is shut down, and a new connection with the agent is set up. The agent is inexperienced, so the CCXML application responds to the request for a call center supervisor and bridges a one-way connection for whispering instructions to the agent, in such a way that the user cannot hear them.

### Advanced enterprise communication services

Combined together, VoiceXML and CCXML are the standards capable of boosting the spread of next-generation communications services while easing the enterprise migration process from old TDM structures to the emerging converged networks. The outstanding advantages that stem from these high-level language standards lie in the flexibility of the applications that can be created through them. These are easily customized and adapted to evolving requirements, they are re-usable and portable to other environments.

- **One corporate number** - to access all personnel and departments, combining all the phone numbers into a single service with intelligent routing that relies on VoiceXML speech dialogues. Customers only ever need to remember one number in order to get in touch with any employee or service within the company.
- **Call screening and filtering** - to allow the customization of call management based on called numbers and user preferences, for example, in order to route every incoming call to a voice box where a caller can leave a message, except a user configurable set of numbers that are routed according to the *find-me follow-me scheme*.
- **Find-me/follow-me** - with parallel or sequential calling so that incoming calls are routed to a user no matter where he roams or what type of device he uses, so he doesn't have to wait by his main phone to receive an important call. Through its ability to manage external events, CCXML can receive presence notifications, which means it is always

aware of the status of employees and can control a status-driven personal greetings dialogue. The system can also automatically announce the user's presence status to the caller.

- **Unified Messaging** - voice and data integration of a variety of services - which leverage UMTS wireless SIP capabilities - to combine voice-oriented interface with data-oriented browsing. Messaging boxes to manage voice mail and e-mail reading, providing hands-free access to the voice mail system. Unanswered incoming calls can be directed by CCXML to a VoiceXML dialogue for email or simple SMS message sending to the intended recipient.
- **Conferencing** - CCXML brings what VoiceXML lacks the most: multi party conference support with moderator and floor control capabilities. The moderator controls who can speak and who should be invited to the conference using voice and DTMF, and the CCXML application can manage announcements indicating when new participants join or leave the conference. VoiceXML dialogues can be added to the conference like any other human participant, they can play the role of "robot moderator" or for the purpose of voice controlling, they can record important issues.
- **Announcing** - a CCXML application can recognize the caller by CLI inspection or invoke a VoiceXML dialogue for the explicit input of credentials and then announce him to the user who can choose between taking the call or sending it to a voice mail application based on another VoiceXML dialogue.
- **Personal assistant** - offering a wide range of services that can be accessed using voice commands, such as dialling entries from a personal address book or the corporate directory, managing and consulting personal appointments with a reminder service that calls back at a given time, mailbox navigation and e-mail reading with functions like reply to, and delete.

Besides the aforementioned present-day features, a number of even more advanced ideas can easily be exploited with the aid of CCXML in order to enrich the Enterprise communication services:

- **Shadow Attendant** - while in conversation, each corporate employee can be temporarily "half-way connected" to a *hot-word* recognition application implementing an "always listening attendant" service which can be activated through a few, simple vocal keys. The attendant will wake up when the "My attendant" utterance is spotted by the hotword VoiceXML application listening on a dedicated CCXML call leg. A number of simple commands can be imagined for everyday tasks, such as: "Attendant, say the number of Mr. White" - when you want to give someone's phone number. The attendant will find it in the user's address book and say the number to its "master" or to the other party, as instructed. Another example is that of a mid-call conference call that can be initiated by saying something like "Attendant, get Mr. Black and Mr. White on a conference call" where Black and White are two of the entries populating the corporate directory or the user's address book.
- **Speaker verification** - a wide range of applications can benefit from speaker verification and identification features when the authentication of the calling parties is required. In order to verify the identity of a user, CCXML can connect the call to two applications simultaneously: one in *full duplex* mode to a VoiceXML 2.0 application, and the other *in half way mode* for verification purposes only. The events will be collected by the CCXML application, which manages the whole process. This feature can be used, for example, to authenticate the users who call into a conference or to grant secure access to the personal assistant by matching a user's utterances with his/her voiceprint, which is stored in the corporate directory system. Alternatively, the caller can be connected to just one VoiceXML 2.1 application capable of recognizing and recording audio concurrently - i.e. while the user claims his identity using his voice, the recorded input is sent to a speaker verification system.

All the aforementioned types of services, which can be effortlessly created using CCXML, require a platform which supports both CCXML, VoiceXML and other relevant standards. The following section of this article will provide a description of how the Loquendo VoxNauta Platform supports the development of such services.

## VoxNAUTA 7.0: A CCXML AND VOICEXML COMPLIANT SPEECH PLATFORM

Due to the outstanding potential of CCXML, the latest release of Loquendo's VoxNauta Platform, VoxNauta 7.0, includes the implementation of a CCXML interpreter. In addition, VoxNauta 7.0 benefits from a completely renewed architecture (shown in Figure 1) which has been designed to fully support all the relevant standards in the speech and call control fields with a pervasive modularity that ensures the highest efficiency. Exploiting a CCXML interpreter frees developers from any dependency associated with proprietary call control implementations while improving their call control application with all the flexibilities and power achievable leveraging the CCXML language.

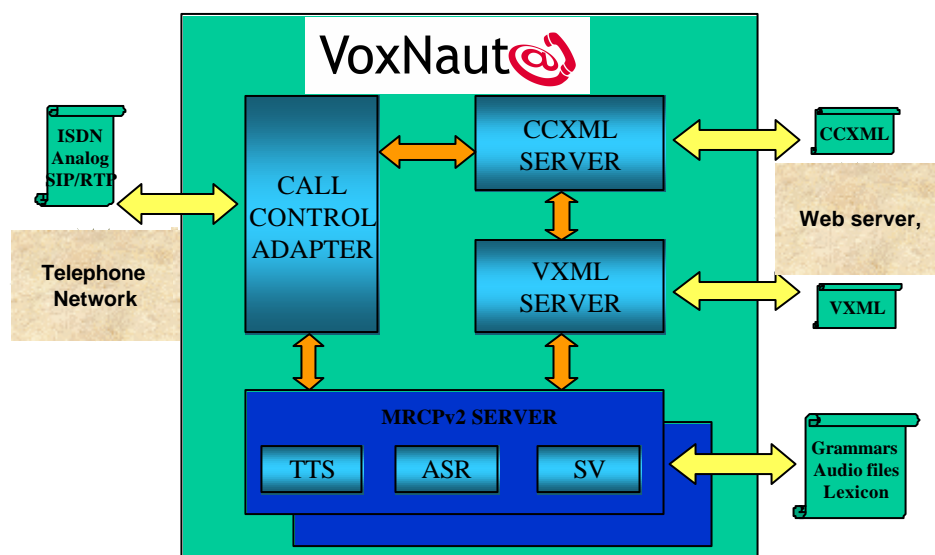


Figure 1 – Loquendo VoxNauta 7.0, renewed architecture

Moreover, combined exploitation of CCXML features and **SIP/RTP** standard VoIP protocol - another protocol exploited in VoxNauta 7.0 interfaces - allows customers to develop a wide range of applications, such as those described above, with dramatic system deployment cost reduction.

In addition to the CCXML, SIP and MRCPv2 standards, VoxNauta 7.0 provides compliance to all high-level, XML based industry standards in the voice area, enabling customers to leverage the strength of Web-based and standard approach for dialog definition, as well as for prompt creation and grammar design.

In specific terms, VoxNauta 7.0 supports:

- **VoiceXML 2.0 [2] and VoiceXML 2.1 [3]:** VoxNauta 7.0 is among the first platforms worldwide to already comply with VoiceXML 2.1 which, with its additional features, allows VoiceXML developers to boost the power of their services adding more flexibility and increasing the efficiency of their VUI dialog design.

- **W3C SSML 1.0** (Speech Synthesis Markup Language) [4]: this markup language provides authors of synthetic speech content with a standard way of controlling aspects such as pronunciation, volume, pitch, rate, etc.
- **W3C SRGS 1.0** (Speech Recognition Grammar Specification) [5] and **SISR 1.0** (Semantic Interpretation for Speech Recognition) [6]: they provide a powerful and standard way to define DTMF and speech recognition grammar syntax and semantics, ensuring the complete portability of services, thus removing any platform vendor legacy.

Moreover, VoxNauta 7.0 exploits the new 7.0 series of Loquendo's Synthetic Speech and Automatic Speech Recognition engines Loquendo TTS and Loquendo ASR, that provide breakthrough synthetic speech quality and important ASR functionalities, including, among others, acoustic model adaptation. The ASR and TTS technologies are integrated through the IETF standardized **MRCP v2** protocol, allowing standard usage of speech technologies and modularity that increase global platform efficiency. VoxNauta 7.0 architecture and packaging guarantees a seamless upgrade of TTS and ASR engines, as well as the effortless addition of new voices and languages as and when they are necessary or become available. This option guarantees that customers immediately benefit from technology enhancements without major platform upgrades, allowing the phased growth of customer installation deployment.

Another relevant enhancement in the new VoxNauta 7.0 design is its inherent operating system independency: from release 7.0 VoxNauta supports the Linux Red Hat 7.3 operating system. The extended support of the Linux Operating System complements VoxNauta's existing support of the Microsoft Windows 2000 and 2003 Operating Systems.

As far as network integration is concerned, Voxnauta 7.0 can be interfaced directly with different PSTN/PBX telephone environments using NMS telephone boards (both digital EuroISDN and analog loopstart), but it also offers a VoIP interconnection option based on SIP-RTP protocols. Due to the **software-only implementation of the SIP-RTP stacks** the VoIP profiles of VoxNauta 7.0 requires no additional third party boards, providing a low cost entry-level, full performing, complete voice platform.

Finally, bearing in mind the impact management tools have on the global quality of the service, VoxNauta 7.0 offers a graphical and easy to use management station that collapses in a single user friendly, hierarchic graphic interface for any configuration, administration or monitoring requirement. The VoxNauta management station can be located on a single VoxNauta server or can be centralized to control complex multiserver installations. A final remark concerns the adoption of a **standard SNMP interface** that gives users the possibility of integrating administration, configuration and monitoring commands within their existing management environment.

## CONCLUSIONS

Despite the continuous, yet significant, advances in speech recognition, a general adoption of automated voice services still depends on the specific service context and on the ability to design the optimal voice user interface, taking service expectations into consideration. Nevertheless, it is undeniable that the development of the various standards in the speech area consolidated in the last couple of years, has significantly boosted the deployment of vocal applications. Loquendo believes it has also provided the basis for a significant revolution in the platform vendor scenario from a competitive point of view. It has determined the obsolescence of proprietary solutions and opened the doors to the conception of low-cost SW-only, interoperable solutions. The CCXML standard is the last step in a process which, by removing the legacy from the proprietary call control mechanism, moves all the aspects of vocal human-machine interface design into the web technology domain .

## REFERENCES:

- [1] RJ Auburn, "Voice Browser Call Control: CCXML Version 1.0", W3C Last Call Working Draft, 29 June 2005.  
See: <http://www.w3.org/TR/ccxml/>
- [2] Scott McGlashan et al., "Voice Extensible Markup Language (VoiceXML) Version 2.0", W3C Recommendation, 16 March 2004.  
See: <http://www.w3.org/TR/voicexml20/>
- [3] Matt Oshry et al., "Voice Extensible Markup Language (VoiceXML) 2.1", W3C Candidate Recommendation, 13 June 2005.  
See: <http://www.w3.org/TR/voicexml21/>
- [4] Daniel C. Burnett et al., "Speech Synthesis Markup Language (SSML) Version 1.0", W3C Recommendation, 7 September 2004.  
See: <http://www.w3.org/TR/speech-synthesis/>
- [5] Andrew Hunt, Scott McGlashan, "Speech Recognition Grammar Specification Version 1.0", W3C Recommendation, 16 March 2004.  
See: <http://www.w3.org/TR/speech-grammar/>
- [6] Luc Van Tichelen, "Semantic Interpretation for Speech Recognition", W3C Last Call Working Draft, 8 November 2004.  
See: <http://www.w3.org/TR/semantic-interpretation/>

## ACRONYMS:

ACD	Automatic Call Distribution
ASR	Automatic Speech Recognition
CLI	Calling Line Identifier
CCXML	Call Control Markup Language
CRM	Customer Relationship Management
DNIS	Dialed Number Identification Service
DTMF	Dual Tone Multi-Frequency
IETF	Internet Engineering Task Force
IVR	Interactive Voice Response
MRCP	Media Resource Control Protocol
PBX	Private Branch Exchange (private telephone switchboard)
PSTN	Public Switched Telephone Network
RTP	Real-Time Transport Protocol
SIP	Session Initiation Protocol
SISR	Semantic Interpretation for Speech Recognition
SMS	Short Message Service
SNMP	Simple Network Management Protocol
SRGS	Speech Recognition Grammar Specification
SSML	Speech Synthesis Markup Language
TDM	Time Division Multiplexing
TTS	Text To Speech
UMTS	Universal Mobile Telecommunications System
Vo-IP	Voice-over-IP
VUI	Voice User Interface
W3C	World Wide Web Consortium
XML	Extensible Markup Language