



# For News Press Two?

VoiceXML and CCXML standards  
offer a better way!

**OptimSys, s.r.o.** tel.: +420 541 143 065  
U Vodárny 2 fax: +420 541 143 066  
61600 Brno info@optimsys.com  
Czech Republic www.optimsys.com

*Imagine that you are calling to a customer service line and an automated voice system greets you instead of obligatory “Welcome, for news press two” with “Welcome, how may I help you”. You answer: “I want to know the latest news” and the system provides you directly with hot news. This is a reality with new generation automated systems that use speech synthesis and speech recognition technologies. The systems are based on VoiceXML and CCXML standards.*

Human communication has been influenced by two major phenomena in the last decade that completely changed it. The first one was a rapidly growing number of mobile phone users, the other one was a massive growth of Internet and the WWW service. Web pages and customer service phone lines became the two most important communication channels with customers and clients for a vast majority of companies and institutions.

While web pages are in most cases a relatively cheap medium with fully automated, continuous operation, the costs of building and operating a call center represent a significant item in the companies' budget. Regardless, it is necessary to support this communication channel because a phone is (unlike Internet) available practically anytime and to anyone.

It is then a logical requirement of the company to automate the telephony communication as much as possible. Automation not only reduces the costs but also brings advantages to the customers, such as shorter waiting time, or the possibility to resolve their requirements fully automatically. This is a great benefit especially in time periods when the line is not serviced by human operators.

The most frequent form of automation are so called IVR (Interactive Voice Response) systems that everybody knows for example from customer care lines of banks and mobile operators. Many markets are still dominated by the previous generation IVR systems where the user has to navigate through a hierarchical system of menus using phone buttons until he/she gets the required information. This information consists of prerecorded pieces of audio that are played back to the user. These systems are characterized by limited flexibility, limited features and vendor lock-in.

## **CCXML + VoiceXML = Voice Browser**

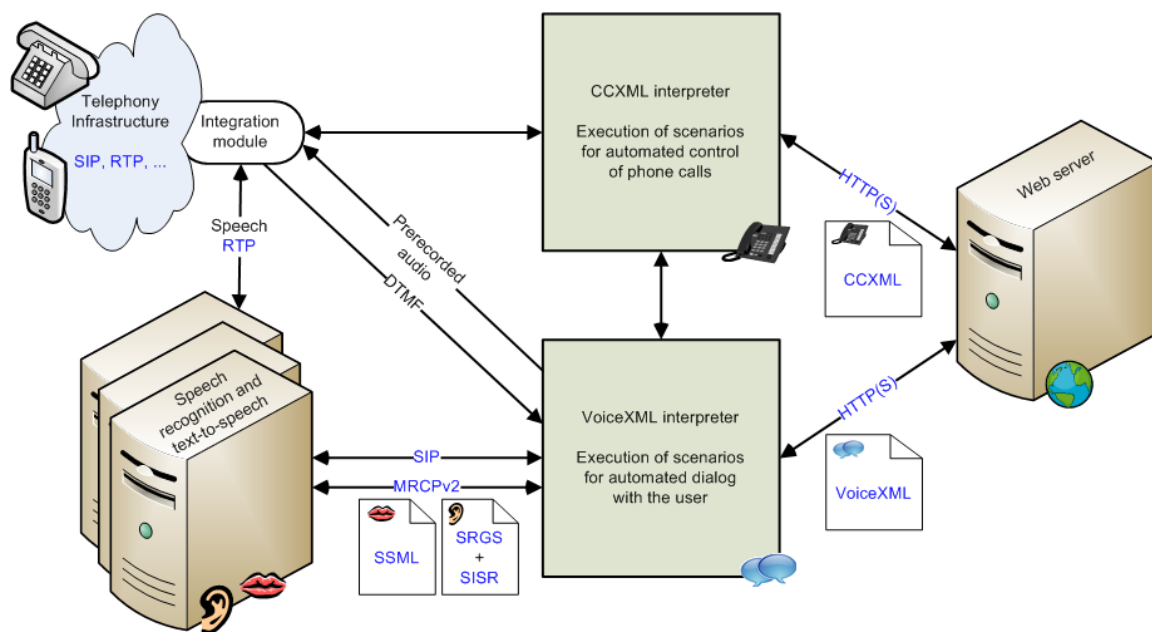
The concept of IVR systems dramatically changed when CCXML (Call Control XML) and VoiceXML standards appeared on the scene. The CCXML language serves to describe scenarios for automated control of phone lines, while the VoiceXML language has been designed to describe scenarios of automated human-computer dialog.

These two standards form together so called *voice browser* that is able to automatically serve the customer. An example of a voice browser is OptimSys' system OptimTalk. Lengthy and unpopular browsing through a hierarchical system of menus using phone buttons might be replaced by speech recognition. The user just says his/her requirement by his/her own words and the computer understands it. The computer can conduct a dialog with the user in case more detailed information is needed. Prerecorded audio can be replaced by speech synthesis, which increases flexibility.

The voice browser scheme is depicted in Figure 1.

The scenarios for automated control of phone lines are described by means of the CCXML language. In this language it is possible to specify in detail:

- incoming call manipulation rules (e.g. answer/reject),
- rules for redirecting calls to another destination (e.g. to an operator),
- rules for attaching calls to conferences,



**Figure 1. Voice Browser Scheme**

- rules for establishing outgoing calls,
- etc.

CCXML scenarios are executed by a computer program called *CCXML interpreter*. The CCXML interpreter is connected with respective telephony infrastructure through an integration module that shields the rest of the voice browser from technical specifics and communication protocols of the telephony infrastructure. In VoIP environments, the integration module typically communicates with the telephony infrastructure using the standardized SIP protocol. Audio streams are then transmitted using the standardized RTP protocol.

One of the operations defined in the CCXML language is connection of a phone call to an automated system that communicates with the user. The respective communication scenario is described by means of the VoiceXML language. In this language it is possible to specify information that should be gathered from the user and describe a way how to gain this information. VoiceXML scenarios are executed by a computer program called *VoiceXML interpreter*.

Among others, the following basic operations can be used within the VoiceXML scenarios:

- using speech recognition for processing spoken user's input,
- processing of DTMF input (input entered by the user using phone buttons),
- recording user's spoken input,
- playback of audio files to the user and
- playback of messages generated from a text by a text-to-speech system (speech synthesis).

The greatest advantage of the VoiceXML language lies in the ability to conduct a dialog with the user and adapt the communication strategy to the current situation. For example, if the user does not tell the computer all the required information, the computer asks for the missing pieces of informa-

tion in next steps. On the other hand, the user has the possibility to tell the computer more information in one step than the user is asked for and thus reduce the conversation length. Also, the sequence of providing the information is not predetermined and may vary.

## Other Standards on the Scene

Another important standard for the voice browser is the Media Resource Control Protocol v2 (MRCPv2). It serves for client–server communication between the VoiceXML interpreter and speech synthesizers and recognizers residing on servers in the network. This distributed architecture ensures desired scalability because speech synthesizers and recognizers are computationally and memory intensive in general.

Again, the RTP protocol is used for transmitting voice to the speech recognizer and from the speech synthesizer. For initialization of a session between the VoiceXML interpreter and the MRCPv2 server the above mentioned SIP protocol is used.

Among other information, the text that should be rendered to speech must be sent to the speech synthesizer. This text may be enriched with a notation that influences the output of the speech synthesizer (for example the voice that should be used, its intonation, speed, placement of breaks, etc.). The Speech Synthesis Markup Language (SSML) has been developed and standardized for this purpose.

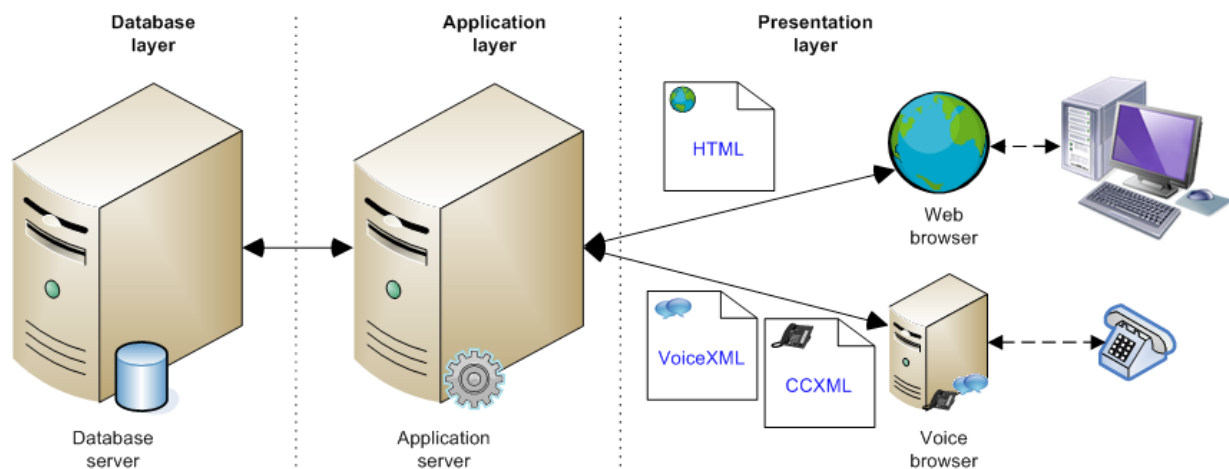
Among other information, the definition of phrases that the user might say must be sent to the speech recognizer in form of so called *grammar*. The grammar format is defined by the Speech Recognition Grammar Specification (SRGS) standard. The grammar can be further enhanced with information describing how to associate a meaning to the uttered phrase because the meaning is the information necessary for holding a dialog. For example, phrases like “Yes,” “Sure” and “Exactly” have usually the same meaning – they represent an agreement. This phrase meaning evaluation process as well as the annotation format within grammars is defined by the Semantic Interpretation for Speech Recognition (SISR) standard.

The CCXML, VoiceXML, SSML, SRGS and SISR specifications are standardized by the W3C consortium and they form together a framework called *W3C Speech Interface Framework*. The SIP, RTP and MRCPv2 specifications are standardized by the IETF organization. All the standards are patent free, with freely accessible specifications. The process of CCXML and MRCPv2 standardization has not been finished yet, however their commercial implementations are quite common, including the OptimSys' implementation.

## Integration of the Voice Browser with Existing Systems

It is obvious that both the above mentioned communication channels, i.e. a web presentation and a customer service line, need to work with the same data and with the same internal logic. Only the form of information presentation is different. The VoiceXML and CCXML technologies have been designed in such a way that it is possible to integrate the voice browser into the standard three-tier architecture that is implemented by a vast majority of modern information systems and web applications. The integration schema is shown in Figure 2.

The user typically sends requests and data to the application server via a web browser and the application server sends responses back to the web browser in form of HTML pages. The web browser then displays these pages to the user. When using the voice browser, the only difference is that the user uses phone instead of keyboard and monitor, the web browser's role is taken over by the voice



*Figure 2. Voice Browser in the three-tier architecture*

browser and responses from the server are sent to the voice browser as CCXML and VoiceXML scenarios. The VoiceXML scenarios then describe the way of how the information is presented to the user in voice.

Integration of the voice browser into the customer's web infrastructure requires no substantial changes in existing systems and applications. The voice browser "only" extends the capabilities of the system with a new communication channel and possibly with new applications.

Communication of the voice browser with the application server runs via the standardized HTTP protocol. The HTTPS protocol can be used to encrypt the communication if needed. Web services can be easily called from VoiceXML and CCXML scenarios via the standardized SOAP protocol. It is also possible to send requests into and receive requests from the CCXML interpreter from any external application via the HTTP protocol.

The voice browser can therefore also serve as a powerful and versatile platform for integration of computer and telephony systems (CTI).

## CCXML and VoiceXML Advantages

Using CCXML, VoiceXML and above mentioned related technologies brings a dramatical cut-down on implementation costs (up to 90 %) and time-to-market (by more than 2/3) [1] in comparison with the previous generation technologies. The solutions are highly flexible, robust, easy to maintain and modify. The orientation on proved web standards allows for a maximum use of existing resources since the web infrastructure is present in virtually every organization nowadays.

Automated IVR transactions also significantly reduce operating costs. According to [2], the price of an automatically served transaction ranges from 0.05 to 0.40 USD, while the price of a transaction handled by an operator ranges from 3.50 to 6.00 USD (data for the U.S. market).

Another advantage of these technologies is the fact that they are based on open, international, patent-free standards. This ensures interoperability between various solution parts provided by different vendors as well as interoperability with technologies developed in the future. Open standards further bring broader possibilities of automation and ensure the existence of wide spectrum of tools from many vendors. They also guarantee a better availability of educated professionals able to cre-

ate and maintain solutions based on these standards. Last but not least, an advantage also lies in a good availability of quality documentation and literature.

Using CCXML and VoiceXML technologies thus reduces the total cost of ownership and represents a significantly lower risks and higher investment protection.

The main differences between the previous and the new generation of automated telephony and voice systems are summarized in Table 1.

<i>Previous generation</i>	<i>New generation</i>
Based on proprietary technologies	Based on open, international, patent-free standards <b>CCXML and VoiceXML</b>
Controlled by phone buttons	Controlled by phone buttons and <b>speech recognition</b>
Playback of prerecorded audio	Playback of prerecorded audio and on-the-fly speech rendering using <b>text-to-speech</b>
Basic Voice over IP (VoIP) support	Seamless <b>integration with Voice over IP (VoIP)</b>
Application development using proprietary tools	Application development <b>similar to design of HTML pages, tools from various vendors</b> available

*Table 1: Comparison of the previous and new generation of automated telephony and voice systems*

## Application areas

Using CCXML and VoiceXML technologies opens entirely new areas for applications accessible over phone. For example, searching a public transport timetable over phone was difficult with the previous generation technologies because typing the location names using phone buttons is very inconvenient. This problem is elegantly solved with speech recognition. Another example might be a system for reading emails that could be realized by speech synthesis.

One of the most common application areas is creation of automated information and reservation lines. Typical are lines providing information about traffic situation, public transport schedules, weather forecast, culture, sports, or tourist information, further airplane ticket, hotel or car reservation lines, goods order and parcel tracking systems, bank and insurance service lines, telecommunication operator customer care lines, or product information lines.

These technologies can bring significant benefits to call centers because they can automatically serve simple, frequently repeated requests, so that the operators can focus on more complicated problems. Different sources state the automation rate from 40 up to 90 %. The voice browser might also be used as a control system for the whole call center telephony, providing algorithms for automated call distribution, predictive dialing etc.

A significant contribution lies in automation of communication during emergency situations such as fire, dangerous substances leakage or other security incidents in objects or areas. The text-to-speech technology together with the technology of automatic control of communication lines makes it possible to automate the communication related items of an emergency plan. Unlike humans, the com-

puter system is able to communicate with all people involved at the same time, which can save a significant amount of time, which may in turn save human lives or reduce economical losses. The computer system also eliminates the risk of human errors made due to the stress. The output for the operator, who can perform other necessary tasks in the meantime, may have a form of a report summarizing the performed communications and their outcome. Reliability of the automated message delivery can be guaranteed by asking the recipient for entering a confirmation code using the phone buttons.

A related area is distant device monitoring and control. Non-critical monitoring systems often send an SMS to given number in case that it is detected that monitored quantities exceed defined limits. Replacing the SMS with a phone call with the possibility of verifying delivery of the message increases the reliability and efficiency of these systems. It is also possible to make a phone number accessible where the system informs the user about current values of monitored quantities by means of speech synthesis. Moreover, the user can be offered the possibility to turn the device on or off, or modify the device configuration over phone. These systems can be deployed in smart houses, data centers and other monitored premises.

The VoiceXML and CCXML technologies are also suitable for creation of user collaboration and unified messaging systems including voice mail systems, reading of emails using text-to-speech, appointment planning and oncoming appointment notifications, automated creation of conference calls, etc.

Another interesting application is an automatic receptionist or switchboard operator. The application might serve for automated switching of incoming and internal calls within the organization. The speech recognition systems are on such level that they can recognize lists counting tens of thousands of names (and a few orders of magnitude more for English and other mainstream languages). The system is also able to conduct a dialog with the user in order to clarify ambiguities. The application can also perform other tasks such as providing general information or taking messages. If the system can connect to an internal attendance system containing information about presence and activities of each employee, the system can automatically decide whether it should redirect an incoming call to the employee's desk, his/her mobile phone, or it should rather record a message.

The voice browser can also serve as a powerful and versatile platform for integration of computer and telephony systems (CTI). A typical example is an integration with customer relationship management (CRM) systems. The CRM system can be easily extended with a "click to dial" feature that uses the voice browser to initiate the call. The voice browser may also notify the CRM system about an incoming call, so that the CRM system can open a screen popup with information about the caller before the user accepts the call.

The potential of the VoiceXML and CCXML technologies has been fully exploited by OptimSys during implementation of the OptimCall system. This system works as an PBX extension and allows the user to set behavior of his/her phone line without the need of changing the PBX configuration. While PBXs often offer only a few basic functions for call processing, OptimCall offers much greater flexibility and detailed settings for each line. Incoming calls can be processed as follows. The call can: ring on the called line (no action); ring on several lines simultaneously, possibly starting ringing on some of them after a delay; be redirected to another number; be redirected to voice mail, recorded message can be sent by email or made available through a web interface; be redirected to an automated voice system (IVR); be ignored or rejected; be recorded when accepted and more. Different operations can be applied to an incoming call based on the current time and phone

number of the caller, and different sets of rules may be applied to working days, weekends, or in case the user is at a meeting, on a business trip or on holiday.

There are many other possible application areas. For example telephone survey systems, help desk extension systems, or systems for an automated distribution of voice notifications and messages, to name a few.

## Are They Really So Good?

The VoiceXML and CCXML technologies are changing the concept of automated telephone and voice systems. They reflect the current market requirements and bring desired flexibility, interoperability and standardization to this area. Thanks to the use of advanced technologies, including text-to-speech and speech recognition, they allow for a higher degree of automation of communication than traditional IVR systems and open completely new application areas towards automation.

According to a Datamonitor research [3], shipments of VoiceXML based IVR systems eclipsed shipments of traditional IVR systems in 2008. According to a T3i Group study [4], 95 % of all shipped IVR systems will support VoiceXML and 90 % of all the systems will use SIP-based VoIP telephony in 2013. These projections clearly show that the use of VoiceXML and CCXML standards for the creation of modern automated telephone and voice applications is the right choice.

## References

- [1] E. Jackson: Speaking Up For Cost Savings In The Call Center: VoiceXML Takes On The Dinosaur Of Legacy IVR. Customer Interaction Solutions Magazine, August 2003.
- [2] Donna Fluss: Why Hosted IVR May Be Right For You. CRMxchange, July 2009.
- [3] Datamonitor: Leading Speech Applications That Will Unlock Enterprise Budgets (Strategic Focus). October 2008.
- [4] T3i Group: InfoTrack for Converged Applications 2008 IVR Market Report. July 2009.