



Using Speech to Extend the Capabilities of the Web 2006

Author: Jonathon Stone, Solutions Consultant, Fluency Voice Technology

Fluency Voice Technology, 3rd Floor, Hythe House, 200 Shepherds Bush Road, London, W6 7NL

Email: info@fluencyvoice.com Tel: 0208 237 6262



Abstract

The World Wide Web has rapidly revolutionised the way we communicate and share information. It has opened up new businesses opportunities and transformed the way existing businesses operate.

The idea of “adding speech to the Web” includes two different scenarios:

- i) Allowing users to control pages using voice commands whilst sitting at their screen and also hearing information played back to them in audible format (extremely useful for the visually impaired)
- ii) Allowing people who either have no web access at all or those who are remote from their PC to access web information and/or perform a broad range of real-time transactions.

This paper reviews the changes taking place in the second of those scenarios. Just as HTML drove the development of the web VoiceXML has emerged as an open standard that has set to transform the telephony industry. Proprietary touch-tone systems of the previous generation which were invariably expensive, slow to deliver and created a situation of supplier lock-in, are rapidly being replaced with solutions based on VoiceXML. This is providing benefits to both customers in terms of consistency and availability of services as well as reduced costs and faster time to market for businesses.

This paper will cover what is driving this change both from the business and customer perspective. It will also highlight the opportunities, benefits and challenges that lie in wait for those looking to transition their web services to the phone.

Introduction

Imagine the scenario, you are driving home and realise you have forgotten to pay an important bill. Usually, you'd use the internet banking service to do this but you're in the car and there's a chance that you will forget once you get home. Instead you can use the banks' 24 x 7 speech-automated telephone service and since it is voice activated it means that you don't have to worry about typing in account details into your key pad. The bill details have already been set up using the internet so it's a straight forward process to select the bill and the amount you need to pay.

The story goes on – you find yourself in a city centre at 7pm suddenly needing a room for the night at a hotel rather than return home. The search begins for a hotel. Again, you think of a well known chain's website and how easy it is to book on line. The same service is available on the phone so you quickly phone up, check availability and book the necessary rooms. You even manage to secure special web rates for your room by using the speech-automated booking service.

Far fetched?

Not at all - these types of services are already in use and are based on integrating the telephone with existing web services.

Everyone has access to a telephone. There are ten times as many telephones in the world as connected PC's. The number of mobile phones continues to increase dramatically and whilst they may facilitate access to the web most people still find it easier and quicker to just talk after all we've been doing that for thousands of years.

There has been a change underway for a number of years to move away from using traditional touchtone interactive voice response (IVR) for providing automated services to using more flexible speech recognition technology. However, the Web is also making a tremendous impact on this transition as it provides ready made access to existing information not only in a graphical format but now also via the telephone in an audible format.

There is now a momentum and an acceptance towards adopting a methodology based on standards such as VoiceXML for delivering speech solutions via the phone. The business benefits are clearly demonstrable as they are based around reduced cost, increased operational flexibility to make changes and updates and less risk of inconsistent or incorrect information being provided to callers.



Why Look to Extend Web to Phone Self-Service? – The Opportunities and Benefits

Today, more than ever, your business needs to work harder, smarter and quicker to retain existing customers and attract new ones. As customers' expectations rise and their propensity to switch suppliers increases your technology needs to be used in ways that provide new, more convenient services that clearly differentiate your business from your competitors.

Another change that seems to be happening is people's way of linking website services to match their needs of receiving information in a format and at time that is convenient to them not just when they are in front of their PC. Website use and terminology is readily understood by most people now and this is influencing their perception of the provision of other services.

However, the PC is not always a convenient way of retrieving information at any particular moment in time, for example if you are out of the office or home all day. Mobile access to the web is increasing through libraries, cybercafés and WiFi but this simply isn't yet prevalent or reliable enough. Not all homes have a web connected PC and some people are simply not comfortable or willing to use a PC. It still takes a few minutes to boot up most PCs whereas the phone is instantly available.

So this is a great opportunity for businesses to reach their customers using speech self-service over the phone in a way they are used to doing on the web. The same infrastructure in place for the web can be used for the telephony channel, the only difference being that the caller uses their voice to request and provide information rather than pointing and clicking a mouse.

Take Travelodge for instance, who have implemented Fluency's speech technology to bring all of the functionality of its advanced Website to the telephone, giving hotel customers the ability to find and book its hotels by phone, 24 hours a day, from any location.

Head of Reservations, Shona Fraser comments "What impressed us the most was the fact that the speech system could use our website as its platform," Shona adds. "We have invested extensively in our website over the last 18 months, and are very proud of some of the innovative features on there, such as the ability to find the best-rate room in the area at the touch of a button. We have been able to maximise this investment by extending the same capabilities to the phone, offering best room rates only previously available on the Web, due to the low cost of handling the enquiry."

Providing a self-serve route to the website through speech recognition is more cost effective than providing a similar service using live agents whose skills can then be applied to more complex services requiring human initiative or up-selling. Many types of live agent calls are becoming impractical due to their cost when compared to web self service, so companies are choosing between increased price (such as premium rate telephone charges) or reduced service – long queues etc. Extending the reach of the Web through speech provides a new solution to this. As new services are added to the Web it is worth considering whether they can also be provided via the telephone as a lot of the infrastructure will already be in place.

The Technology

XML – short for eXtensible Mark-up Language – is a flexible and structured way to define documents and information. It is a non-proprietary, vendor-neutral standard maintained by the World Wide Web Consortium (W3C) the group that develops common protocols for ensuring interoperability across the web. XML is similar to HTML, the mark-up language used to define web pages however it defines what the data is rather than what it should look like. It can easily be extended to create other documents and information standards and is ideal for the exchange of data between applications.

VoiceXML is a simple tagged language that is similar to HTML but much smaller in size with only 50 elements. Whereas HTML allows users to navigate a website with clicks of a mouse button with VoiceXML the caller simply uses their voice (or where appropriate touch tone) to control the flow of the application. VoiceXML acts as an interpreter between the telephone and the website content and can easily be installed alongside existing web logic.



One of the other great benefits of VoiceXML is its portability. Applications developed to the W3C specification will be portable to other platforms thereby removing the dependence on proprietary development languages and development interfaces that influenced the IVR market for so long. Since VoiceXML is similar to HTML it means there is an existing mass of web developers ready to start developing using the standard.

“Web Services” refers to a new set of standards that define a modular application model for Web applications. Web services are application components with XML-based interfaces and provide a flexible way to deploy new application modules with existing legacy applications.

Web services are based on the following standards:

- Web Services Definition Language (WSDL) – defines the capabilities of the Web Service, where it resides and how it is invoked.
- Simple Object Access Protocol (SOAP) – describes the message exchange mechanism for Web Services.
- Universal Description Discovery and Integration (UDDI) – provides the mechanism for clients to find other Web Services.

VoiceXML can easily integrate to Web Services which means applications can be developed quicker and cheaper as the integration to the back end data is likely to already be in place. Previously, proprietary IVR systems depended upon significant effort to integrate directly to backend systems particularly involving extra effort by IT departments which would often lead to delays and extra costs.

You may also hear mention of SALT, Speech Applications Language Tags, another subset of XML. SALT has been heavily endorsed by Microsoft but recently they have also started to support VoiceXML which makes some people think VoiceXML will eventually be the dominant language. There are differences between the two languages as SALT is aimed at being multimodal in that it supports multiple interfaces within the same call (whereas VoiceXML really only supports Voice or DTMF input and pre-recorded or synthesised speech as output). This means it is possible to use different input and output methods with devices such as PDA's.

One weakness of both VoiceXML and SALT is their support for call control in that they support only basic call transfers. This is currently overcome by using either CCXML (Call Control eXtensible Markup Language) or CTI software. CCXML can provide the call management, event processing and conferencing capabilities currently missing in VoiceXML and SALT.

The Importance of Design

Callers interact with a speech application through a telephone with a voice user interface (VUI) analogous to the way they would access a website through the screen via a graphical user interface (GUI). There are significant differences between the VUI and the GUI and it isn't always a straightforward process to replicate the GUI in the form of a VUI.

In terms of design, a poorly designed website will frustrate a user but they are still likely to be able use it whereas a poorly designed VUI will frustrate callers to the extent that they will look to exit to an operator or simply hang-up.

VUI design is thus important because it is often a customer's first contact with a business and will therefore influence their perception of that business. Well designed applications can increase customer satisfaction by providing timely and efficient access to information and transaction processing. Telephones are readily available and require no additional accessories or devices to be used.

However, designing successful VUI is dependent upon two key issues (a) understanding basic human capabilities and (b) understanding the caller's needs and goals in relation to the business goals.



There are limits to the human cognitive processing which means we struggle when we are presented with too much information. Providing services via the telephone will place extra loading on callers' memory and there is a lack of persistence with the information i.e. the information can only be played once to the caller unlike the web where it can remain on the screen.

Consideration also needs to be given to the fact that callers may be easily distracted whilst using a telephone service and there may also be a significant amount of background noise. Whilst the caller interaction also suffers from being sequential and slow compared to other media clever design can help to overcome these "limitations".

There is a significant amount of evidence to support the claim that callers still prefer speech to touch tone applications. This is for a number of reasons:

- i) Speech is natural – well designed prompts elicit natural answers making applications easy to use
- ii) Speech is intuitive and efficient – it's quicker and easier than pressing digits on a telephone key pad
- iii) Speech can be appealing – well crafted applications can be entertaining as well as efficient
- iv) Hands and eye free – essential if driving for example.

Designing a successful VUI is both a science and an art. The science involves using speech recognition technology effectively and the art involves designing a caller experience that is compelling, efficient and satisfying.

Getting It Right First Time

The main thing that will affect the challenge you face in getting an application "right first time" is whether you opt for a Packaged Application or a custom build. There is another strong analogy with the web here. If you go back to the first few years of commercial web sites, each one was built from scratch by teams of HTML designers and Java coders, with subsequent change management involving manual redevelopment. These days, all but the most specialist of organisations use packaged modules and content management tools. The speech market has only recently made this change, with the first Packaged Applications emerging in the last two years. If you are able to find a suitable Packaged Application then you can limit the challenges to the following;

i) Configuration of the dialogue to match your business processes - You will need to consider changing existing processes to make them more natural to complete using speech, trading off the cost of these changes with improved application performance.

ii) Persona selection - Careful selection of a "persona" whose voice characteristics match the brand values of the business helps the caller match their initial mental model to that they will experience in the call. Properly matched this will provide the caller with a more effective and rewarding experience coupled with an associated increase in task completion.

iii) Web Integration to allow access to back end systems - The use of software stubs can help speed up the development and testing processes. The code pretends to be an interface to another remote machine that is perhaps more complicated or not currently available and again code developed for the original web site may well be reusable for the telephony service.

If you are unable to find a Packaged Application which can be used as the basis for your application, you face a number of other challenges as you will have to design, develop and test the whole application, including the VUI. This will require significant external involvement from specialist service providers, with the associated time and cost impact. This becomes apparent particularly in the initial launch phase where a pilot period will be essential to "tune" the custom built application using live calls to flush out the inevitable design issues around prompts, grammars and so on. Whilst this can add several months and significant cost it is essential in order to get your custom application performing to the level where it can support a high quality customer experience.



Once an application is into business as usual operation it is not the end of the project. Websites are dynamic as will be your speech self-service system. Inevitable variations to business processes or launches of new products or services often result in the need for call flows changes and the recording of new prompts. It is important that these changes can be effected quickly to maintain the highest possible levels of customer service. This is where you will especially see the benefits of the kind of content management tools that should be part of a Packaged Application. Without these you will be dependent on third party service providers to make the changes to your application - think back to the maintenance challenges of those websites projects of the 1990s!

Conclusion

As the World Wide Web continues to impact on our everyday lives so our desire for anytime, anywhere access to information forces businesses to consider more effective use of other channels to provide access to existing Web information and processes.

It makes sense for businesses to consider using speech to extend the reach of the Web for a number of reasons. It is now feasible to reuse code from existing Web applications to provide telephony services that can be automated with a speech interface. Integrating an automated speech application with an existing Web infrastructure is now quicker and cheaper thanks to the emergence of new standards such as VoiceXML. This provides a faster and more feasible ROI as well protecting existing Web investment.