A multimodal E-commerce application coupling HTML and VoiceXML

E-commerce - User interface and interactions - Web applications

José Rouillard
Laboratoire TRIGONE, CUEEP
Bat B6 - Université des Sciences et Technologies de Lille
59655 Villeneuve d'Ascq cedex - France
Tel: +33 (0)3.20.43.32.65
Fax: +33 (0)3.20.43.32.79
jose.rouillard@univ-lille1.fr

ABSTRACT

This paper describes our experiences in developing a prototype of electronic commerce application using both HTML and VoiceXML languages. This virtual shop allows a customer to use her voice in input and see the results of her request on her PC screen, but also to interact on a traditional web form, and hear on her phone the effects of this action.

Keywords

E-commerce, VoiceXML, Multimodality, Personalization

1. INTRODUCTION

It is clear that HTML is more useful for textual and graphic presentation while VoiceXML was especially conceived for typically non-graphic interactions: telephone-based applications. *"Considering that the technologies that multi-modal applications will be based on are still at very early stage in their evolutionary process, it will be a long while before we see any device that will allow us to browse in voice and text modes at the same time."* [1].

Providing a multimodal E-commerce application through heterogeneous devices (phone and Web Browser for example) is not so easy because this kind of interactions are not supported by the same languages. *"The most important drawback that we found in our attempts to nevertheless build a multimodal architecture around VoiceXML, is that it is not possible to make an active voice dialog running in a VoiceXML browser aware of events that occur outside the voice browser, e.g. at a visual interface: VoiceXML neither allows for linking in Java applets that could receive pushed notifications, nor does it provide any other interface for external events* [3].

1.1 Using HTML and VoiceXML within the same application

The Figure 1, gives a synoptic vision of the dynamic application files, in which we can see the course a user should follow in order to buy clothes and others goods, and to add it in her virtual basket. Both, visual and phone-based interactions are used in this man-machine dialog.

The voice interaction begins with the *welcome.vxml* file. If the user answers "yes" at the question "Are you already one of our clients ?", the next called file is *customer.asp*, which continue to interact in oral mode, generating a VoiceXML document.

In the other case, the machine synthesises a vocal message, and gives a specific number to the user. This number is the client number that the customer will give thanks to the graphic interface, in a HTML page, dynamically created (see Figure 2). Creating HTML or VoiceXML pages is an ordinary job. But there are not supposed to be used in the same time. The main problem we have to face, here, is a synchronisation one.
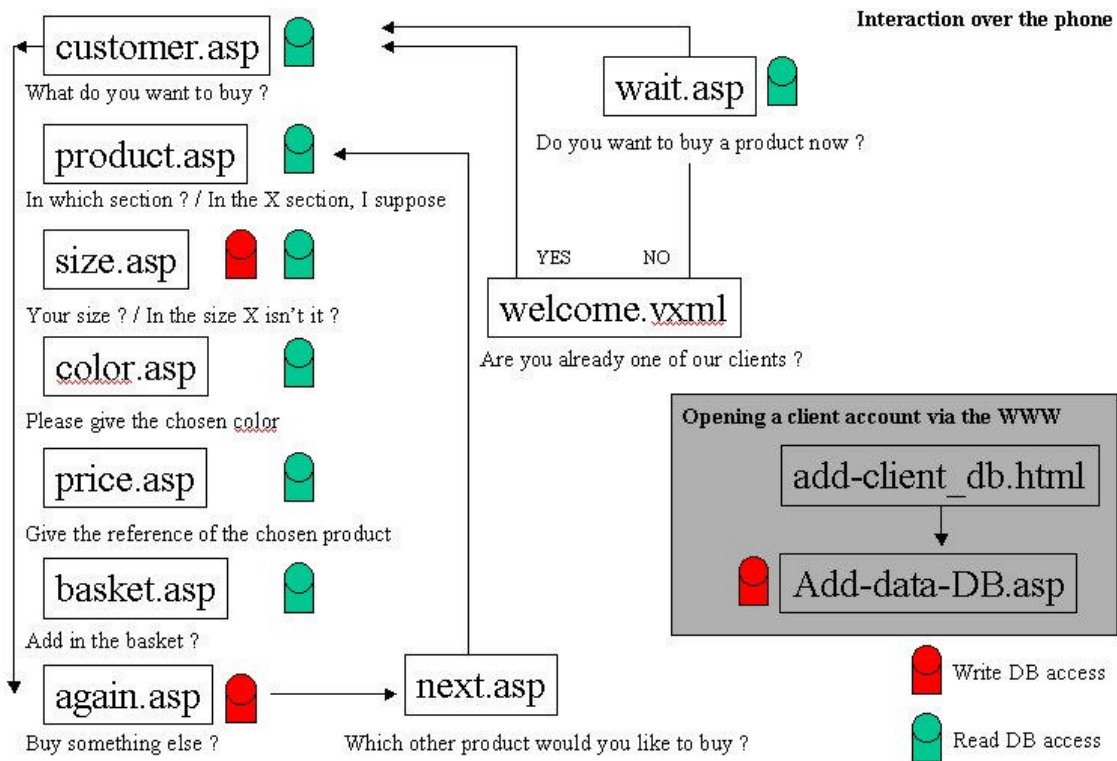

Figure 1: Path followed by the consumer on our commerce site

1.2 Synchronisation issue

"*One of the biggest problems in the way of multi-modal progression is synchronisation.*" [1] If a speech recognition system can more or less recognize what you said (it knows what you can say because it knows a particular grammar and vocabulary), it is very difficult, nay impossible, to catch what you said if the linguistic domain is not limited. That's why we decided to propose to the new clients to identify themselves in a traditional web page, as we can see on Figure 2.

Figure 2: The information given in this web page will be used by the voice browser

From this moment, the information given by the user on the web page are usable in the entire application, including the voice interaction side. This is possible with dynamic grammars.

The Figure 3 shows an example of oral dialog between Miss Jennifer Hutchinson (which was unknown a few minutes ago) and the machine. In the oral interaction, the machine asks the user to go to a traditional web page and to complete a form (see Figure 2). From this moment, the synchronization process takes place, and the oral side is waiting for the data coming from the textual and graphic side. If the registration number used by the consumer is correct, the interaction continues on an oral mode.

```
10:32:15 C: Welcome on our web site. Are you already one of our client ?
10:32:26 H: no
10:32:26 C: We are going to create your client account. In order to
register yourself, please give in the Web page value 39 as client number.
Did you correctly hear your client number ?
10:32:41 H: yes
10:32:41 H: yes
10:32:41 C: Ok. I'm waiting while you are completing the page.
10:32:47 C: Your registration is correct Miss Jennifer Hutchinson. Would
you like to buy something now ?
10:32:56 H: yes
10:32:56 C: What do you want to buy Miss Jennifer Hutchinson?
10:33:06 H: some dresses (…)
```

Figure 3: The oral dialog uses information coming from the textual Web page

In the same way, synchronization in output is possible between graphical and vocal interfaces for the same application. The Figure 4 below shows how a traditional web page gives the results of an oral user's request. This page is automatically refreshed every 5 seconds (for example), and retrieved the current data useful for the user to choose a product in the virtual shop. So this multimodal interaction allows to see on the screen some information that the vocal browser cannot display, like the photos of the goods, in our example.

Référence Client	Titre	Prénom	Nom	Date de naissance	Taille	Pointure
14	Monsieur	Edouard	Martinot	23/12/63	40	41

Référence Produit	Description	Rayon	Photo	Prix en Frs	Couleur	Taille
3	Pantalon Charpentier	Homme		188	orange	40
35	Pantalon gris	Homme		188	gris	40
36	Pantalon a pinces	Homme		200	blanc	42
39	Pantalon velours	Homme		278	mauve	44
43	Pantalon bleu	Homme		254	bleu	40

Figure 4: The web page gives automatically the results of an oral request

2. PERSONALIZATION

There are many ways to adapt and personalize the data we have to present to the users. It can be done, by observation, according to the user's preferences, membership to a group of user, etc. Those preferences can be asked to the user directly, or automatically studied and stored, each time a task is completed. The information could be used for the current session or for the next ones. We have decides to establish two kind of personalization : one based on logic elements (deduction), and the other on observations.

2.1 Personalization based on logic elements

An intelligent system is supposed to help by providing relevant information to the users. An example of the machine's anticipation is given in the dialog of the Figure 5 : knowing that this customer is a man (Mr. John Smith), it suggests the men's section, in order to find the pants. Of course, this proposition can be rejected, if needed, in the cases where the user is not buying something for herself/himself. That's why the system only suggests information. In the Figure 5, we can see how the consumer rejects the machine's proposition, and changes from "man" to "sport" section, during the dialog.

```
(...)
18:00:40 C: Glad to hear you again Mr. John Smith. What do you want to buy
?
18:00:46 H: pants
18:00:46 C: There are 13 trousers available. I suppose that you want to buy
this article in the men's section.
18:00:55 H: no
18:00:55 C: In this case, choose another section Mr. Smith.
18:00:59 H: sport
(...)
```

Figure 5: Example of personalization : the client rejects the computer's suggestion

2.2 Personalization based on observations

Another way to use relevant personalization is to observe what the user is doing and to try to anticipate what she would probably do or ask for. For example, in the Figure 6, the machine has already asked (and stored), in a previous dialogue the size of the client. This time, it suggests the size that seems to be correct.

```
(...)
14:02:25 C: Which other product would you like ?
14:02:30 H: a dress
14:02:41 C: You would like this article in the size 36, isn't it?
14:02:48 H: yes
14:02:49 C: There are 7 women dresses, with the size 36. Please give me the
desire color.
(...)
```

Figure 6: Using personal information in the dialog

As we can see on the Figure 7, at the beginning of the dialog, the size of this client is unknown in her profile's file (the French words "Taille : inconnue" in the first table, can be translated as "Size : unknown"). So the machine asks the client's size in order to help to choose a suitable trousers for the consumer. In a future transaction, the machine will suggest the size or the section according to the user profile and collected data.

Figure 7: Personalization process during the man-machine dialog

3. CONCLUSION AND FUTURE WORKS

Different approaches are possible to give the user multimodal interactions feasibility. We have shown that, even if it's not easy, because of synchronization matters, it is possible to create a multimodal application that allows the user to act on a particular device, and to see or hear the results of her interaction on another one, within the same application.

Multimodality is often studied on a same device. The contribution of this works is to demonstrate that multimodality can be implemented through heterogeneous devices and heterogeneous languages. We have shown that HTML and VoiceXML can be coupled, thanks to synchronisation mechanisms. Dynamic languages such as ASP or PHP can be used to post and retrieve data coming from different devices.

We use personalization processes to adapt the presented information to the needs of the users, according to their profiles or based on observations. The personalization adds to our prototype a kind of "intelligent" comportment, by suggesting some relevant information.

4. REFERENCES

1. Breitenbach, S., Burd, T., Chidambaram, N., Astrid Andersson, E., Tang, X., Houle, P., Newsome, D., Zhu, X.,Early Adopter VoiceXML, Wrox, 2001.
2. Lucas, B., VoiceXML for Web-based distributed conversational applications, Communications of the ACM, vol. 43, N°9, 2000.
3. Niklfeld, G., Finan, R., Pucher, M., Architecture for adaptive multimodal dialog systems based on VoiceXML, Eurospeech, 2001.