NRC Publications Archive
Archives des publications du CNRC

Enabling rapid development of multimodal data entry applications
Kondratova, Irina; Durling, Scott

This publication could be one of several versions: author’s original, accepted manuscript or the publisher’s version. / La version de cette publication peut être l’une des suivantes : la version prépublication de l’auteur, la version acceptée du manuscrit ou la version de l’éditeur.

Publisher’s version / Version de l'éditeur:

Conference Proceedings On the Move to Meaningful Internet Systems 2007, 2007

NRC Publications Archive Record / Notice des Archives des publications du CNRC :
https://nrc-publications.canada.ca/eng/view/object/?id=ffeb9bc5-44b0-4e0f-9795-e87343402524
https://publications-cnrc.canada.ca/fra/voir/objet/?id=ffeb9bc5-44b0-4e0f-9795-e87343402524

Access and use of this website and the material on it are subject to the Terms and Conditions set forth at
https://nrc-publications.canada.ca/eng/copyright
READ THESE TERMS AND CONDITIONS CAREFULLY BEFORE USING THIS WEBSITE.

L’accès à ce site Web et l’utilisation de son contenu sont assujettis aux conditions présentées dans le site
https://publications-cnrc.canada.ca/fra/droits
LISEZ CES CONDITIONS ATTENTIVEMENT AVANT D’UTILISER CE SITE WEB.

Questions? Contact the NRC Publications Archive team at PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca. If you wish to email the authors directly, please see the first page of the publication for their contact information.

Vous avez des questions? Nous pouvons vous aider. Pour communiquer directement avec un auteur, consultez la première page de la revue dans laquelle son article a été publié afin de trouver ses coordonnées. Si vous n’arrivez pas à les repérer, communiquez avec nous à PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca.

National Research
Council Canada

Conseil national
de recherches Canada

Institute for
Information Technology

Institut de technologie
de l'information

NRC · CNRC

*Enabling Rapid Development of Multimodal Data Entry Applications **

Kondratova, I., and Durling, S.
November 2007

* published in the Conference Proceedings On the Move to Meaningful Internet Systems 2007: OTM 2007 Workshops. Vilamoura, Portugal. November 25-30, 2007. Lecture Notes in Computer Science, R. Meersman, Z. Tari, P. Herrero et al. (Eds.): OTM 2007 Ws, Part I, LNCS 4805, pp. 377–386, 2007, © Springer-Verlag Berlin Heidelberg 2007. NRC-49853.

Copyright 2007 by
National Research Council of Canada

Permission is granted to quote short excerpts and to reproduce figures and tables from this report, provided that the source of such material is fully acknowledged.

Canada

Enabling Rapid Development of Multimodal Data Entry Applications

Irina Kondratova[1] and Scott Durling[1]

[1] National Research Council Canada Institute for Information Technology
46 Dineen Drive, Fredericton, NB, Canada E3B 9W4
{Irina.Kondratova, Scott.Durling @nrc-cnrc.gc.ca }

Abstract. Evaluations of mobile multimodal applications are frequently conducted in the laboratory which allows recreating the intended usage context, whilst tightly controlling and monitoring the testing environment and tasks performed. For this type of technology evaluations, in order to speed up the development process for speech-enabled multimodal applications, we developed a multimodal authoring tool. This rapid prototyping tool allows us to quickly produce a proof of concept multimodal data entry application that we can utilize to evaluate the feasibility and usability of using speech in a particular usage context, e.g. on the construction field, for municipal inspections, for gaming applications, etc. Our tool makes the development of a simple multimodal data entry prototype application easy even for a novice user. This paper describes design considerations for the multimodal prototyping tool and elaborates on our experience working with the tool.

Keywords: multimodal, speech recognition, user evaluations.

1 Introduction

As new modalities for human computer interaction emerge, in-depth testing is required to validate their efficacy for use in a variety of application domains. One such modality that has long been envisioned as the gold standard of human computer interaction is speech. Thus far, the potential of speech interaction in computing has not yet been fully realized due to low accuracy rates of existing speech recognition engines, especially in noisy conditions. As speech technology continues to improve and evolve, performance and usability evaluations will play a key role in facilitating wide spread acceptance of speech as an interaction modality for mainstream use.

Research suggests that combining modalities - that is, the pairing of interaction techniques such as gestures, handwriting, stylus, and speech, with varying forms of feedback (e.g., visual, audio, etc.) - may prove to be the most effective means by which to make mobile applications usable [1]. Speech-based multimodal interfaces allow integration of visual and speech interfaces through the delivery of combined graphics and speech, on handheld devices [2]. This enables more complete information communication and supports effective decision-making. It also helps to overcome the limitations imposed by the small screen of mobile devices. Most of the PDAs and Pocket PCs today have a small screen size of about 3.5”x 5”, and mobile phones have even smaller screens [3]. The small screen size, and the need to use a pen to enter data and commands, presents an inconvenience for field users - especially if their hands are busy using other field equipment, or instruments.

In spite of the great promise of speech technology, due to inherent difficulties of speech processing in the field environment [4], it is advisable for designers of speech applications to conduct empirical studies to understand the reasons behind successes and limitations of their applications, and to investigate alternatives for speech in a particular context [5]. As a result, effective tools for authoring multimodal interfaces are now required in order to develop and evaluate the use of multimodality for mobile users.

Mobile data collection can benefit greatly from the freedom found when given more choices for interaction with a mobile computing device. This paper discusses the tool we developed to rapidly create multimodal (in our case, speech and keyboard/mouse/stylus) applications for data entry scenarios. Following some background information on technologies for coupling speech with other modalities, we will discuss multimodal authoring tools developed by others and present general considerations we used to design our tool. Next, we will describe the major design blocks of our authoring application and discuss an example of the application developed with our prototyping tool. We conclude with the summary of our experience working with the tool and opportunities for future research work.

2 Combining speech with other modalities

There are two different models for implementing multimodal interaction on mobile devices. The fat client model employs embedded speech recognition on the mobile device and allows local speech processing. This model can be only implemented on a handheld computing device that has enough processing power, for example on a laptop PC, tablet PC, or a Pocket PC (e.g. HP IPAQ). The thin client model places speech processing on a portal server and is suitable for mobile phones.

One of combinations of the markup languages that are currently used for developing multimodal web applications is VoiceXML + XHTML (X+V) [6]. It combines XHTML and a subset of VoiceXML (Voice Extensible Markup Language). VoiceXML provides an easy, standardized format for building speech-based applications. Together, XHTML and VoiceXML (X+V) enable Web developers to add voice input and output to traditional, graphically based Web pages. This allows the development of multimodal applications for mobile devices based on the fat client model that includes a multimodal browser and embedded speech recognition on a mobile device, and a web application server. As a part of our research focus on investigating speech as an alternative interaction modality, specially for mobile users, we developed several speech applications based on VoiceXML technology and multimodal applications utilizing X+V technology [7, 8]. By choosing VoiceXML we were able to work with existing multimodal browsers (Opera^{TM} or NetFront^{TM})

that incorporate IBM ViaVoice© speaker-independent speech recognition engine. For the user evaluation purposes we found it more appropriate to use the speaker-independent speech recognition engine as opposed to a speaker-dependent speech engine that need to be trained by the users (such as Dragon Naturally Speaking). It is important to note, that the accuracy of speech recognition is influenced by the level of environmental noise and by the quality and type of the microphone used. In our user evaluations of the speech data entry in the construction field we tested speech input for mobile data entry under three different levels of background noise – 70dB-80dB, 80dB-90dB, and 90dB-100dB. Our results showed that noise level significantly affected the accuracy of data entry and suggested that there may even be a threshold of approximately 80dB noise level beyond which the accuracy achievable with speech input may prove unusable [4, 9].

3 Authoring for multimodality

With the steady increase in availability of devices that can utilize speech recognition and other modalities, the task of authoring multimodal interfaces has started to gather significant attention of the multimodal and mobile computing research community [10-12]. Researchers found that rather than developing each modality independently for each target platform, efficiency can be improved by utilizing design tools to facilitate the development of multimodal interfaces.

Our literature review revealed that there are several schools of thought regarding the tools for authoring of multimodal interfaces. Some tools are envisioned as truly design tools while model-based tools simply model the interface and hide the design aspects in presentation generation [12]. Design tools allow the designer to have more control on the final product, whereas model-based tools generate static interfaces based on presentation generation engines.

Currently researchers have utilized three types of tools to aid in development of multimodal interfaces including: add-on toolkits, graphical designer WYSIWYG tools, and model based development tools. Add-on toolkits are built to add the needed functionality to a pre-existing interface design tool, as is the case with IBM's Multimodal Toolkit [13] which utilizes the IBM Rational Development Environment. This type of a multimodal authoring tool is by and large useful for learning to program using X+V multimodal markup language, but does not accelerate the development of multimodal interfaces, nor enable the process to be performed by a non-technical user.

On the other hand, development tools based on rich client platforms are utilized to facilitate the development of multimodal interfaces. As an example, the MONA project utilized the Eclipse development platform to build a design tool for their MONA presentation server [12]. This server, upon a request, translates a proprietary XML format which defines a multimodal interface and tailors it to the specific device type. This system allows developers to have a great deal of control over the design of graphical and speech components.

Multimodal Teresa is another tool developed for making multimodal interfaces [10]. Utilizing a model, this authoring tool allows navigation and data entry tasks to be defined in an abstract form first, and, following this, particular variable names and other options can be specified for each modality. Finally, multimodal web pages are generated by the tool based on user's specifications. A shortcoming of this approach is that the model based tools: "…traditionally suffer from lack of control over the generated result by the developer" [12].

In our research studies on using speech as input modality for mobile data entry in the field, we conducted user evaluations of usability and efficiency of using speech versus stylus [4] as well as evaluated the efficiency of different types of microphones for mobile speech data entry [9]. Based on this experience, we found that the process of developing multimodal prototypes for our evaluations was quite lengthy, required specialized programming experience with VoiceXML and XHTML, and familiarity with speech recognition applications and speech grammars.

To assist in user evaluations of multimodal applications and to speed up the prototype development process for multimodal applications, we developed our own tool for multimodal authoring called 'Multimodal Maker'. When designing the tool, we adopted a model based approach, similar to what was used to develop "Multimodal Teresa". The next section of our paper describes design considerations for the Multimodal Maker tool.

4 Design considerations for the tool

Our goal was to create a multimodal authoring tool for development of *prototype* multimodal data entry applications. First of all, we wanted to shorten the development time; from our past experience we learned that the development of such a test application could take up to a year. We also wanted to have a tool that the average user, including a user with only basic computer programming skills, will be able to utilize to design a multimodal application in a few minutes. This multimodal application will be subsequently utilized in conjunction with the Multimodal Web Browser (Opera^{TM} or NetFront^{TM}) to evaluate speech and text data entry in different conditions (e.g. high level of noise, mobile use, nautical environment, etc.).

Since out goal was to *automate* the development of multimodal applications for novice users, we have chosen a model based approach for designing our tool and abstracted away the specifics of the interface design. As it was mentioned previously, model based approach in building multimodal applications makes modifications rather difficult. In the Multimodal Maker tool the design aspects are automated after the model has been created and, therefore, not controlled by the average user. We felt, nonetheless, that the ease of use achieved by utilizing this simple approach is meeting our needs.

In addition to creating a tool for authoring the multimodal xml (mxml) files, our goal was to create a complete data entry application. To achieve this, within the authoring tool we built a module for database creation. This module automatically creates a local database and which allows saving the data collected by the multimodal data entry application. The database functionality allows us to capture some valuable information collected during user evaluations, such as speech recognition factors, user inputs, etc. for subsequent detailed analysis of the multimodal interaction. Finally, by utilizing IBM DB2 Everyplace©, we can easily extract our data into a spreadsheet application for further analysis.

Data entry tasks can typically be reduced to a list of data one wishes to capture. Thus, we represented our data entry tasks as an entity object called Form. Forms naturally contain sections and grammars that could be used by input fields in the form. Input fields are divided into section objects that encapsulate a grouping of related input items. Speech recognition grammars in our application are stored on a form by form basis, and the functionality is provided for a designer to import grammars from previously defined forms. Grammar objects are conceptually reduced to lists of information items (Choices), and are used to populate data input types such as list boxes, drop down boxes and radio buttons. The simple hierarchical object design diagram for our tool is presented in Fig. 1.

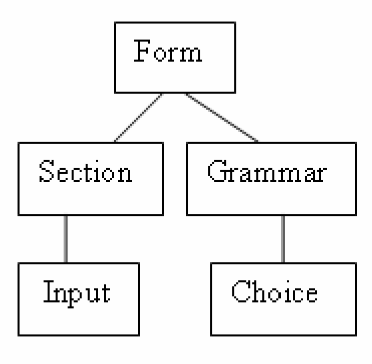


Fig. 1. Object Design Diagram

In addition to the grammar specification functionality mentioned previously, all built-in grammars for speech data entry provided by the X+V standard are available within our authoring tool. Addition of specialized grammars is easily accomplished as well, but requires knowledge of the java speech grammar format (JSGF). For example, for prototype evaluation purposes, we have created our own JSGF grammar for decimal numbers and added it to the tool's built-in grammars. A summary of multimodal input types and their grammar requirements is presented in Table 1.

In our development process, the processing modules for generation of XHTML + VoiceXML (X+V), the database, and their connecting infrastructure were built utilizing the factory design pattern. The default generator produces code for a multimodal web browser running on a Windows XP based system. The design pattern we have chosen allows for adding factory objects for various other platforms/modalities in the future.

Table 1. Input Types and Grammar Requirements

Type	Format	User Specified Grammar Required
Single Select	Drop Down	Yes
	Radio Button	
	Yes No Check Box	No
Multiple Select	List Box	Yes
	Check Boxes	
Text	Number Digits Time Phone Number Date Currency Decimal	No

5 Using the authoring tool

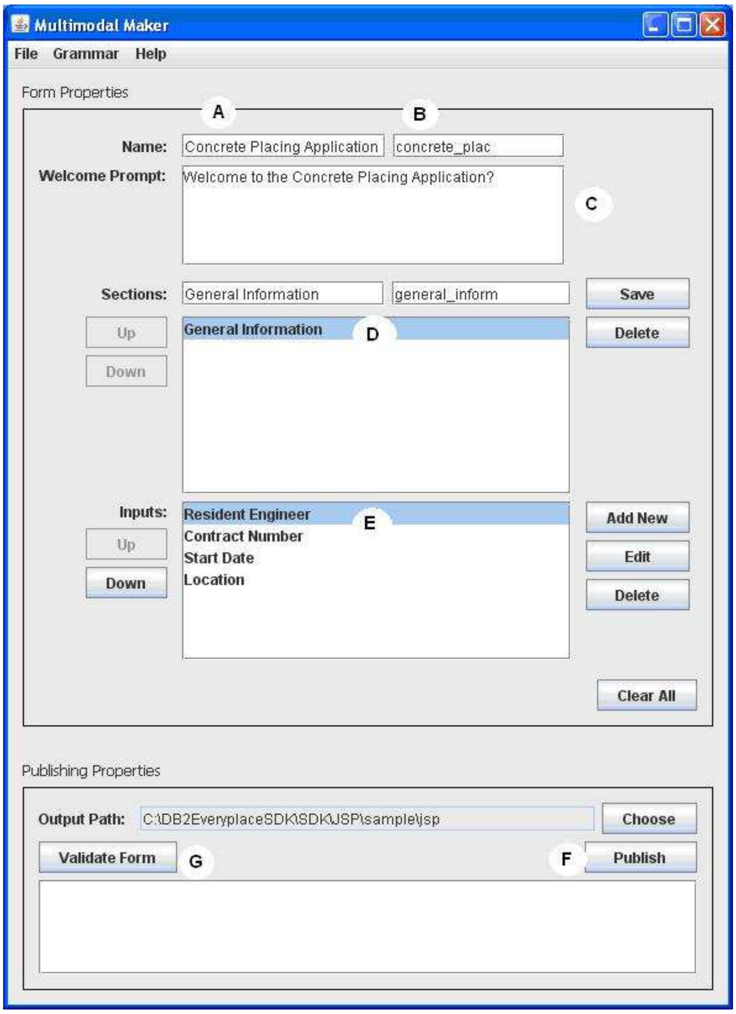
The current version of the tool generates a multimodal interface similar to a prototype application we created previously for field concrete inspection data collection [7]. In our previous study, in order to test the usability of using either stylus or speech for data collection, a limited subset of the input data fields from the original application was used in a controlled lab-based evaluation [4]. Such a prototype application could be easily created with our authoring tool.

Annotated screen shot below (Fig. 2) presents the steps of creating a multimodal data entry prototype. Each part of the interface has a display name and an abbreviation. The display name (A) is used for speech and visual display purposes, and the abbreviation (B) must not contain spaces or special characters so that it can be used for programming of the interface components. Keeping the abbreviation name editable by the designer was implemented in order to facilitate easy troubleshooting and greater control over naming of components.

To design a form, the user enters the display name. Following this, a suggested abbreviation will appear automatically (B). Sections of the application could be added by entering a section name and an abbreviation (D). Sections order can be changed using the up and down buttons. When the user chooses a section, the inputs for this particular section are presented in the input box (E).

The user can specify a welcome prompt upon opening the multimodal application. Inputs are defined by entering a display name, an abbreviation, a type of the input and a grammar (if applicable), as presented in Fig. 3 (A). User defined grammars are reduced to lists of data, making their definition simple, as seen from Fig. 3 (B). A validation feature is available for the Grammar and Form definition screens that automatically checks for the proper formatting of abbreviation names.

The generation factory is invoked when the user chooses ‘publish’ upon specifying an output path at the bottom of the main screen (See Fig. 2). A default corporate style is applied to the multimodal application as per Fig. 4. Users can then enter data in the fields of the form using speech, keyboard/stylus or a combination of speech and keyboard/stylus.



Name: Concrete Placing Application | concrete_plac

Welcome Prompt: Welcome to the Concrete Placing Application?

Sections: General Information | general_inform

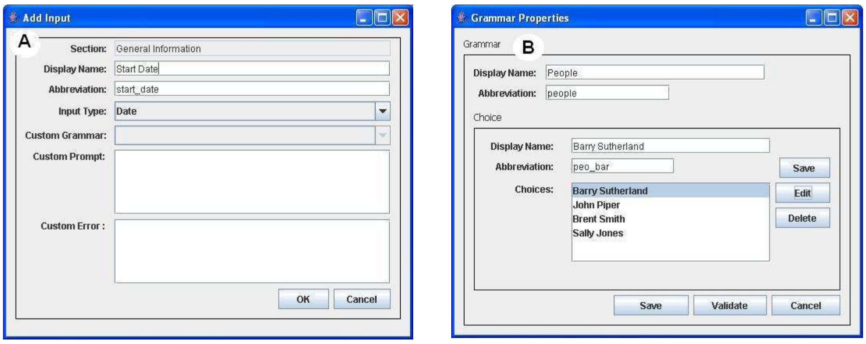
General Information

Inputs:
Resident Engineer
Contract Number
Start Date
Location

Output Path: C:\DB2EveryplaceSDK\SDK\JSP\sample\jsp

Fig. 2. Data Entry Form Designer

As a result of our development efforts, we developed a tool that allows us to rapidly create prototype multimodal applications we can use to carry out user evaluations of the multimodal mobile technology. In addition, several modules of our authoring tool have been packaged into reusable components for easy integration into other multimodal application generation projects.

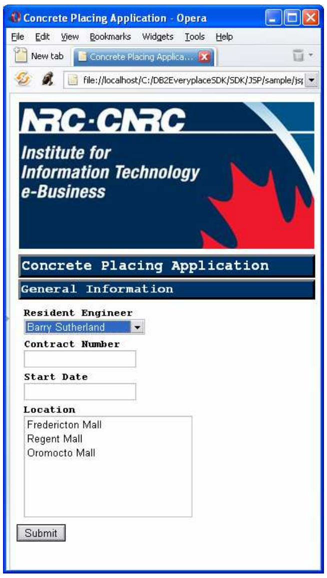


Section: General Information
Display Name: Start Date
Abbreviation: start_date
Input Type: Date
Custom Grammar:
Custom Prompt:
Custom Error :

Grammar
Display Name: People
Abbreviation: people

Choice
Display Name: Barry Sutherland
Abbreviation: peo_bar
Choices: Barry Sutherland
John Piper
Brent Smith
Sally Jones

Fig. 3. Specifying Inputs and User-Defined Grammars

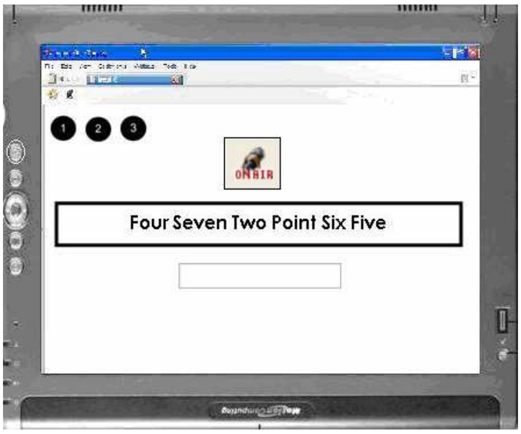


Concrete Placing Application - Opera
File Edit View Bookmarks Widgets Tools Help
New tab
Concrete Placing Applica...
file://localhost/C:/DB2EveryplaceSDK/SDK/JSP/sample/jsp
NRC·CNRC
Institute for
Information Technology
e-Business

Concrete Placing Application
General Information
Resident Engineer: Barry Sutherland
Contract Number
Start Date
Location: Fredericton Mall
Regent Mall
Oromocto Mall

Fig. 4. Example of Resulting Multimodal Interface

As an example, during the design and planning stage of the research study on testing the efficacy of microphones for speech-based data entry [9] we realized that the standard data entry form (similar to one presented in Fig. 4) was not appropriate for our evaluation purposes. As an alternative, we utilized reusable modules of the Multimodal Maker tool to rapidly generate an application that prompts for one user input per web page as shown in Fig. 5. The test application was generated programmatically utilizing Multimodal Maker tool and accompanying methodologies, with the exception of some added user interface functionality such as the number of input attempts we had to program manually. By utilizing the Multimodal Maker tool to develop the test prototype and data collection tools, we significantly shortened the development time and efficiently collected data during user evaluations.



Four Seven Two Point Six Five

Fig. 5. Microphone Testing Application

6 Conclusions

As a result of this research project, we developed an authoring tool that enables rapid generation of prototype multimodal applications for user evaluations of multimodal technology in the laboratory and in the field. We found that by utilizing this authoring tool we can reduce the time required to develop a multimodal application for speech/stylus/keyboard data entry to a couple of hours by a novice user as compared to months previously required for development of the same application by an experienced programmer. Another advantage in developing the tool is that now we are able to provide this authoring tool to our clients who want to evaluate the feasibility of using speech date entry in a particular context, but do not have in-house expertise with multimodal development.

The Multimodal Maker authoring tool currently has limited functionality and does not allow for incorporation of mixed initiative dialog capabilities as well as graphics and widgets to enhance the functionality of the multimodal user interface developed. This will be a subject of our future work to further broaden capabilities of the tool.

References

1. Oviatt, S., *Taming Recognition Errors with a Multimodal Interface.* Communications of the ACM, 2000. 43(9): p. 45-51.
2. Hjelm, J., *Designing Wireless Information Services",.* 2000: Wiley Computer Publishing, John Wiley & Sons, Inc.
3. Pham, B., Wong, O. *Handheld Devices for Applications Using Dynamic Multimedia Data*. in *2nd International Conference on Computer Graphics*

and Interactive Techniques (GRAPHITE). 2004, June 15-18. Singapore: ACM Press.

4. Lumsden, J., Kondratova, I., Langton, N. . *Bringing a Construction Site Into the Lab: A Context-Relevant Lab-Based Evaluation of a Multimodal Mobile Application* in *Proceedings of the 1st International Workshop on Multimodal and Pervasive Services (MAPS 2006)*. 2006, June 29. Lyon, France.
5. Schneider, T.W., Balci, O. *VTQuest: A Voice-based Multimodal Web-based Software System for Maps and Directions*. in *44th Annual Southeast Regional Conference*. 2006, 10 - 12 March. Melbourne, Florida: ACM Press.
6. Axelsson, J., Cross, C., Lie, H.W. , McCobb, G. , Raman, T. V., Wilson, L. *XHTML+Voice Profile 1.0*. 2001 [cited 2007 July 12]; Available from: http://www.w3.org/TR/xhtml+voice/.
7. Kondratova, I. *Speech-enabled Handheld Computing For Fieldwork*. in *International Conference on Computing in Civil Engineering(ICCC'2005)*. 2005, July 12-15. Cancun, Mexico.
8. Kondratova, I., Goldfarb, I. *M-learning: Overcoming the Usability Challenges of Mobile Devices*. in *International Conference on Networking, International Conference on Systems and International Conference on Mobile Communications and Learning Technologies (ICN/ICONS/MCL'06)*. 2006, 23-29 April. Mauritius: IEEE Computer Society Press.
9. Lumsden, J., Kondratova, I., Durling, S. *Microphone Efficacy for Facilitation of Speech-based Data Entry (In Press)*. in *British HCI '07*. 2007. Lancaster, England.
10. Paternò, F., Giammarino, F. *Authoring Interfaces with Combined Use of Graphics and Voice for both Stationary and Mobile Devices*. in *Working Conference on Advanced Visual Interfaces (AVI'2006)*. 2006, 23-26 May. Venezia, Italy: ACM Press.
11. Polymenakos, L.C., Soldatos, J. K., *Multimodal web applications: design issues and an implementation framework.* Int. J. Web Engineering and Technology, 2005. **2**(1): p. 97-116.
12. Simon, R., Wegscheider, F., Tolar, K. *Tool-Supported Single Authoring for Device Independence and Multimodality*. in *7th International Conference on Human Computer Interaction with Mobile Devices & Services (MobileHCI'05)*. 2005, 19-22 September Salzburg, Austria: ACM Press.
13. Miranti, R., Jaramillo, D., Ativanichayaphong, S., White, M., *Developing Multimodal Applications using XHTML+Voice*. 2003.