# Multimodal application for foreign language teaching

Teresa Magal-Royo[1], José Luis Giménez-López[1], Blas Pairy[1], Jesús García-Laborda[2], Jimena Gonzalez-del Rio[1]

(1) Department of Graphic Engineering
Universidad Politécnica de Valencia,
Valencia, Spain
e-mail tmagal@degi.upv.es

(2) Department of Modern Philology
Office Virginia Woolf
Universidad de Alcala
28801 Alcala de Henares
Madrid, Spain
e-mail jesus.garcialaborda@uah.es

***Abstract*— The current development of educational applications for language learning has experienced a qualitative change in the criteria of interaction between users and devices due to the technological advances of input and output data through keyboard, mouse, stylus, tactile screen, etc. The multiple interactions generated in a natural way by humans during ordinary communication can be transferred in a sequential way to devices like PDAs, PC Tablet, etc. depending on the users' needs to carry out specific tasks that allow humans to adapt to their nearest learning context. This paper shows the possibility of establishing multimodal architectures within the applications for specific language learning areas with ubiquitous devices, evidencing the technical and formal aspects necessary for their accomplishment that are currently being developed at the Universidad Politécnica de Valencia (Spain).**

***Keywords: Teaching English as a foreign language, TEFL, Multimodal user interfaces, MUI, foreign languages, teaching, Computer Aided Learning Language,CALL, usability, Mobile Assisted Language Learning, MALL.***

## I. INTRODUCTION

Research and applications related to Computer Aided Language Learning, CALL [1], have today enabled different methods for online learning to be developed for Teaching English as a Foreign Language, TEFL [2]. The most significant advances have been made by adapting the applied technology to conventional language teaching and in the creation of applications involving online environments and using the web that have given greater independence to the end user when learning a language in a tailor-made way.

In fact, one new emerging area of research is in the use of mobile devices with technology adapted to on-line environments for language learning that enables both teacher and student to have an environment allowing them to teach and learn at any time.

Mobile Assisted Language Learning (MALL) describes an approach to language learning that is assisted or enhanced through the use of a handheld mobile device [3], [4].

MALL is an area of research into Mobile Learning (m-learning) and Computer-Assisted Language Learning (CALL) that involves the development of user-oriented applications for learning or teaching languages, taking into account the latest technological advances, the need for user interaction and the specific kind of learning aimed at Second Language Acquisition, SLA [5].

It is precisely these fundamental criteria that have enabled us to develop a multimodal application for smartphones that is intended to improve access to digital multimedia content and to help the user learn by guiding them towards the acquisition of language skills such as reading comprehension, grammar, composition, oral comprehension and so on.

One of the great contributions made by the concept of multimodality is the ability to create adaptable human-machine communication environments based on the use of different means of data input and output that allow the user to switch the means of interaction according to their social and physiological needs. In fact, mobile devices allow for this concept by integrating various technologies that enable the user to switch the use of voice, stylus or keyboard depending on their communication needs.

Using multimodality allows users to find a more accessible and usable environment, as it allows for adaptability to the environment taking into account the user's cognitive abilities or limitations.

On designing the application that has been created, the choice of sequential multimodality [6] was taken into account, following the guidelines established by the World Wide Web Consortium, W3C, to validate a digital environment through the use of two types of interaction (touch and voice).

Two key aspects for the technical feasibility of the project were also assessed, relating to:

- Previous studies regarding user preferences as regards multimodal environments.
- The study of the reduction or minimization of errors by the user when entering data in such environments.

### *A. User preferences for carrying out tasks*

In general, the early studies related to multimodality and the use of multiple communication means by the user

concentrated on validating the dual methods of interaction with the user based on their preferences when carrying out specific tasks or using specific data input devices. Early empirical studies focused on assessing user preferences when using digital maps or drawing programs.

Oviatt conducted research on the multimodal or unimodal use of an interface based on dynamic localization interactive maps [7], [8] demonstrating the feasibility of the input data (specific choice within a digital map), using speech and an electronic stylus by means of an automatic simulation technique [9].

The experiment's design was based on three types of interaction: interaction by voice only, interaction with the stylus only, and multimodal interaction using speech and stylus. The results of Oviatt's work over those years determined the user's preferences in two aspects:

- The preference for the stylus to locate and draw shapes on a map.
- The preference for speech or voice mode for requesting information about areas marked on the map, labels already established on the map, and the use of descriptive commands.

Cohen [10] established an empirical study by comparing a direct manipulation interface called ExInit with a multimodal system based on the use of a stylus and voice, called QuickSet [11].

In terms of digital environments for on-line language learning, we can say that there is no empirical study available to determine the benefits in preferences for using one communication system over another with a mobile device. Thus, the use of both a tactile screen and voice were taken as the examples to follow for interactively moving through the learning tests and for progress through the different tests to be carried out.

In the final application, the use of one system over another should be chosen by the user before performing the test, with the touch method of interaction being the preferred system because this way the user establishes a prior mental adaptation as to how they are going to input the requested data or what method of navigation they are going to use.

Dual pre-selection is considered essential for the future validation of future multimodal systems, taking into account the specific characteristics of each test to be performed. In fact, an official language exam consists of several sections whose purpose is to validate the knowledge acquired by performing specific tests in each area of learning. Therefore, there is a need to determine what the most suitable multimodal pairs might be in future in order to carry out a complete exam.

### *B. Reducing mistakes by the user on entering data*

Multimodal interfaces can reduce mistakes made by users as they can enter and confirm data in several ways. In fact, the technique related to this concept is called "cross-mode compensation" [12].

The "cross-mode compensation" system demonstrates that combining inputs using different modalities can improve recognition and the performance of a specific task. If multimodal integration can work with a distribution of possible inputs for each input mode, early recognition may help direct the search for the correct end result.

In some cases, when recognition using the two types of interaction is wrong, multimodal integration allows mistakes to be avoided (mutual compensation).

Oviatt [13] performed the first assessments of compensation by using a multimodal system based on stylus movements and voice in the QuickSet system for use in interactive digital maps. This defined a unified multimodal integration model, "Unification-based multimodal integration" [14], which enabled the specific types of structures to be defined that would improve on mistakes when performing tasks.

Multimodal integration based on specific or finite situations is applied directly to applications with established information routes. Research concerning problem-solving in speech recognition and the use of a stylus within an order confirmation service [15] showed that:

- The speech system is more acceptable and has fewer mistakes than the keypad writing system.
- Users prefer flexibility of interaction when correcting mistakes they have made.

Rudnicky and Hauptmann [16] conducted several studies related to the design principles of multimodal interfaces for the correction of speech compared to data input via a conventional keypad. The tasks to be carried out focused on entering a number, correcting it and confirming the numbers provided. The three methods of interaction were: interaction via voice, keypad and multimodal voice and keypad.

The results obtained showed the speech mode of interaction to be the fastest and the one with least mistakes when providing specific numbers or values.

As regards language learning, this aspect has been considered to be essential in specific language-learning tasks through speech, since correcting pronunciation can help the user improve their language diction.

As for the application created, the use of speech has led to a second data input channel that favours, on the one hand, the performance of specific tasks involving choosing the correct answer, validation of a task, choosing a task, etc., and on the other hand the use of a method of progression through each task by recognising specific words.

## II. DESIGN OF THE ENVIRONMENT

The environment created was based on the traditional design for an English language exam that tests knowledge for university entrance in Spain and which is currently done on paper with a number of sections related to reading comprehension, grammar, etc. The contents are created each year and are validated by experts in the field of languages at each university.

By using this type of design, it was easier to adapt the content on paper to the digital environment through the use of multimedia technology for viewing videos or accessing sound files. So, in general the following tasks were performed:

- Adaptation of the contents written on paper to the digital medium, taking into account the type of task to be performed.
- Adaptation to a sequential navigation structure based on a breakdown of the interactive tests, following a predetermined route.
- Adaptation of digital content to the medium (small screen, small font size, etc.)
- Adaptation of the programming language to enable multimodal sequencing of the two chosen means of communication (touch and speech) for solving tasks.

The programming language for data management and storage was Java, since the three main principles on which Java programming is based are: the use of object-oriented programming (OO), the ability to run the same program on several different platforms or operating systems, and compatibility with the Android operating system on which this kind of application is based in mobile phones.

The visual design of the screens took into account the technical and formal accessibility criteria set by W3C for online digital environments, including any restrictions established for mobile devices or small screens. These technical criteria helped to improve and interpret the end flow of information desired, as well as to limit the number of screens and accesses to the exam's sections.

Two devices were chosen for the testing: an HTC Desire and a Samsung i900 Galaxy-S, both using the Android 2.1 operating system.


Figure 1. Multimodal environment for language learning

## III. VALIDATION OF THE ENVIRONMENT CREATED

The first validation tests created for the environment were set up in an emulator on a computer with expert users who helped to identify shortcomings in the interaction mechanism created and to experience the two options created. The experts selected were involved in the field of language teaching, giving language classes in the Polytechnic University of Valencia's Language Centre, as well as computer experts who helped to debug the problems caused by cross-platform compatibility.

The validation phase with students will be carried out soon, using an assessment test of the qualitative level of acceptance in conducting this type of testing on mobile devices, and more particularly for conducting official access exams to the University in the area of languages. Criteria related to usability, functionality and the level of learning satisfaction in using this environment will enable this type of environment to be validated in future.

## IV. CONCLUSIONS

The preliminary conclusions of this paper focus on demonstrating the feasibility of using multimodal environments for testing or examination by teachers and students in the field of language learning.

The use of multimodal environments will extend and improve the levels of accessibility and usability in the web via mobile phones. This will enable new learning methods to be established based on the use of next generation mobile devices.

Future work in this research shall therefore concentrate on assessing the impact of the application on the end users: language learners. At the same time, the technical feasibility of using it in the classroom shall be studied, as well as how language teachers can use mobile devices to create tests and tasks in online, digital environments that help students gain independence in learning at any time, anywhere by using a mobile device.

## ACKNOWLEDGMENTS

This work was carried out through research conducted in the project: "Analysis and verification of adaptation and accessible multimodal interaction for language examinations on mobile devices". It has been funded by the Universitat Politècnica de València as part of the program: 'Projects for new lines in multidisciplinary research – PAID-05-10'.

## REFERENCES


[1] Levy M. (1997) CALL: context and conceptualisation, Oxford: Oxford University Press.

[2] Celce-Murcia M. (2001).Teaching English as a Second or Foreign Language. Heinle ELT, 3rd edition.

[3] Chinnery G. (2006) "Going to the MALL: Mobile Assisted Language Learning", Language Learning & Technology 10, 1: 9-16, [Online]: http://llt.msu.edu/vol10num1/emerging/default.html

[4] Shield L. & Kukulska-Hulme A. (eds.) (2008) Special edition of ReCALL (20, 3: 2008) on Mobile Assisted Language Learning: http://www.eurocall-languages.org/recall/r_contents.html#sep08

[5] Krashen, S. (1982). Principles and Practice in Second Language Acquisition. Pergamon Press.

[6] Multimodal Interaction Requirements. The World Wide Web Consortium W3C. [Online]: http://www.w3.org/TR/mmi-reqs

[7] Oviatt S. (1996). Multimodal interfaces for dynamic interactive maps. CHI. Human Computer Interaction. Pp 95-102.

[8] Oviatt S. (1997). Multimodal interactive maps: Designing for human performance. Human Computer Interaction Pp. 93-129.

[9] Oviatt, S., Cohen, Ph., Fong, M, and Frank, M. (1992): "A rapid semi-automatic simulation technique for investigating interactive speech and handwriting". Second International Conference on Spoken Language Processing. ICSLP´92, Pp.1351-1354.

[10] Cohen Ph. Johnston M., McGee D., Oviatt S., Clow J., Smith I. (1998). The Efficiency of Multimodal Interaction: A Case Study. ICSLP. 5th International Conference on Spoken Language Processing. Sydney, Australia.

[11] Cohen, P.R., Johnston, M., McGee, D., Oviatt, S., Pittman, J., Smith, I., Chen, L., and Clow, J. (1997) QuickSet: Multimodal interaction for distributed applications. In Proceedings of the Fifth ACM International Multimedia Conference ACM, NY, 31–40.

[12] Rhyne, J. R., & Wolf, C. G. (1993). Recognition based user interfaces. In H. R. Hartson & D. Hix (Eds.), Advances in Human-Computer Interaction, 4, 191-250.

[13] Oviatt, S. (1999). Mutual disambiguation of recognition errors in a multimodal architecture. In CHI '99. ACM Press, New York.Pp 576-583.

[14] Johnston, M., Cohen, P.R., McGee, D., Oviatt, S.L., Pittman, J.A., and Smith, I. (1997). Unification-based multimodal integration. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics, ACL*. pp. 281–288.

[15] Oviatt and van Gent (1996). Error resolution during multimodal human-computer interaction. Second International Conference on Spoken Language Processing, ICSLP´96 .

[16] Rudnicky and Hauptmann (1990). “A Comparison of Speech and Typed Input” HLT ´90. Proceedings of the workshop on Speech and Natural Language. Pp. 219-224.