

# Speak up your mind: Using speech to capture innovative ideas on interactive surfaces.

Oswaldo Huerta, J. Alfredo Sánchez, Salvador Fuentes, Ofelia Cervantes

Laboratory of Interactive and Cooperative Technologies

Universidad de las Américas Puebla, México

{osvaldohg, j.alfredo.sanchez, salvador.fga, ofelia.cervantes}@gmail.com

## ABSTRACT

This paper presents Vozze, a user interface that has been designed to speed up the process of recording and managing ideas and their associations during innovation activities. Vozze integrates three major concepts: multi-tactile surfaces, speech recognition, and workspaces that are dynamically created and linked to each other. Our main goal is to provide the advantages of using a technological platform in innovation processes while maintaining a natural interaction. Multi-tactile surfaces are based on simple hand gestures we have designed. Speech recognition addresses issues related with text entry on interactive surfaces, which typically relies on cumbersome virtual keyboards. Finally, dynamic workspaces allow users to easily represent and share their ideas, as well as to create associations among them. We describe a prototypical implementation of Vozze as well as results from initial usability and user experience evaluations.

## Keywords

Interactive surfaces, multi-tactile interfaces, voice interfaces, speech recognition, innovation.

## INTRODUCTION

Teamwork has become one of the main ways to organize group tasks with the purpose to achieve common goals. While performing group activities it is often necessary to plan, share, discuss, and propose new ideas. Innovation seeks radical changes that often result from group sessions in which ideas are generated, discussed and refined. Although traditionally these generally unstructured activities rely on conventional whiteboards, flipcharts and markers, various technologies are being used to support this work, and they increasingly include interactive surfaces such as multi-touch tables, screens or whiteboards.

Interactive surfaces offer evident advantages for presentations and information management tasks in which users do not have to introduce new data or newly generated

ideas.

Thus, it has become common, for example, to watch TV newscasts in which multi-tactile screens are used to interactively present charts, maps and other graphical representations at different levels of detail. Multi-touch screens also are being used widely for advertising and entertainment purposes, but rarely for any tasks in which end users enter data.

We are particularly interested in supporting innovation activities by introducing the use of multi-touch surfaces in early stages of the innovation process. Tasks related to adding newly created information are difficult to perform because, as of yet, interactive surfaces do not provide natural and efficient ways to enter information. Typical means to insert new data on large interactive surfaces are virtual keyboards that are displayed on the screen, or physical keyboards. The use of these devices cause users to divert their attention from the objects on the screen and the ideas being discussed to the virtual or physical mechanism required for data entry. Moreover, it is hard to capture all the information or ideas generated during collaborative sessions by typing them into a computer or by taking notes.

Once ideas are captured, another challenge is to manipulate and represent the inter-relationships among those ideas. Natural and efficient gestures as well as means for managing workspaces are needed.

We have developed Vozze, a system concept designed to provide a natural solution in scenarios where interactive surfaces do not provide complete support for collaboration and innovation processes. Based on multi-touch surfaces, Vozze uses voice as a natural way to capture ideas, specific gestures for manipulating and editing their textual and graphical representations, and dynamically linked workspaces to organize ideas and their inter-relationships. We find it important that voice is perhaps the most natural way we have to express our ideas, thoughts and feelings, hence our emphasis on making this form of expression the main focus of this project.

In the remainder of the paper we discuss work that is related to our initiative, as well as the main concepts considered in the design of Vozze. We then describe a prototypical implementation that integrates existing and

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

10<sup>th</sup> Brazilian Symposium on Human Factors in Computer Systems & 5<sup>th</sup> Latin American Conference on Human-Computer Interaction. IHC+CLIHIC 2011 October 25-28, 2011, Porto de Galinhas, PE, Brazil.

Copyright 2011 SBC. ISSN 2178-7697

new hardware and software components. Next, we discuss results of a user study conducted with users that performed innovation tasks around a multi-tactile table. Finally, we present some of the issues addressed by ongoing and future work, as well as some conclusions derived from the work conducted thus far.

## RELATED WORK

This section reviews briefly work that is related to voice-based interaction. We also refer to work that motivated our approach to dynamically linked workspaces. We do not delve into details of speech or gesture recognition, as these components of *Vozze* mostly integrate work that has been described elsewhere.

Speech recognition and synthesis technologies provide a natural interaction alternative for many computer-supported tasks. Speech allows users to communicate with computers naturally using natural, verbal language, thus requiring very little training. Although these technologies have existed for several decades, they have become widely usable relatively recently [15]. Nowadays, there exist important advances in speech technologies that provide high quality levels of recognition and synthesis services. These services can easily be integrated in collaborative environments for expediting data entry and producing automatic responses. Frequently, their functionality is made available through libraries or programming interfaces that facilitate the implementation of new applications [4], [10], [9], [11], [14].

Many interesting examples show the potential of using speech technologies for facilitating interaction. The *VoiceCode* project, for example, allows users to develop tools for programming by voice [16]. *Speech dasher*, is a novel interface for entering text using a combination of speech and navigation via a pointing device (such as a mouse). In *Speech dasher* a speech recognizer provides the initial guess of the user's desired text while a navigation-based interface allows the user to confirm and correct the recognizer's output [19].

In recent years, various alternatives to classical input devices such as keyboard and mouse as well as novel interaction paradigms have been proposed. Particularly, multimodal interactions have been developed to overcome the limits of each input channel taken alone. Thus, [2] [3] describe a system based on the integration of speech and gaze-based input for interaction with a real desktop environment.

Interactive surfaces can support innovation and collaborative tasks. Recently, there has been a growing interest in how large shared displays can be designed to facilitate small co-located group working, including the use of interactive wall displays and tabletop displays to support playing games, selecting and viewing images or sorting information [7] [12]. The synchronous collaboration environment termed “Strata” [13] supports interaction and

collaboration among local and remote participants who share a workspace and communicate by pointing and annotating on objects.

In order to enrich and accelerate the innovation process, voice can be incorporated as a mechanism to interact with tactile surfaces. There are various efforts to change regular applications to applications where the user is able to use interactive surfaces by applying multi-touch gestures and using speech [17] [18]. In those applications, voice is used to dictate predefined commands to the application. This kind of effort has been applied to software such as the voice search feature in Google Earth for mobile devices and World of Warcraft [20], including gestures to delimit areas or trace routes. In those applications it is possible to use speech, but only for dictating selected commands or receiving answers produced with the voice of predefined characters. Free speech recognition capabilities are not available yet.

*NiCE* [6] is a platform that proposes a technological solution for work meetings. *NiCE* includes interaction with multi-touch surfaces, personal computers and regular paper sheets. *NiCE* offers a complete solution for collaborative work in the office, supporting interaction with interactive whiteboards where users can change colors, add short annotations and highlight information. A special kind of markers is used to integrate the information written into regular paper. Although *NiCE* integrates paper and digital media to support group meetings, *NiCE* still does not incorporate speech to capture the ideas generated by the participants. The contribution of *NiCE* is to provide different ways of interaction but still does not allow for the inclusion of new ideas created in real time, since the only way users have to include them is to type ideas using a computer or writing the complete idea using a marker.

Innovation processes require collaborative work with spontaneous contributions that are easier to be expressed using speech instead of writing. Traditional techniques for adding ideas are usually slow and may break the creative nature of the innovation process. The lack of speech recognition is a limit to the interaction with the surfaces. *Vozze* proposes to include voice as a method to capture new ideas easily and naturally.

An effort that has been influential for our work has been *KMS* (Knowledge Management System). *KMS* is a multi-user spatial hypermedia system developed several decades ago [1]. It models very closely the way associative thinking works and emphasizes rapid system response to allow for users to focus on the ideas they intend to capture. Central to *KMS* is a model based on the concept of frame, which is a unit that records items related to an idea, and that can be linked to other frames. *KMS* was originally developed in C and Pascal, and ran on Unix machines. Although certainly other models have emerged more recently, we refer to *KMS* due to its simplicity, which makes it appropriate for interactive surfaces that are expected to be intuitive and

straightforward. Moreover, it is worth mentioning that a multi-platform, Java-based version of KMS called *Expeditee* is currently being developed at the University of Waikato, New Zealand.

**VOZZE CONCEPTS**

During meetings, participants are used to take notes, sketch diagrams or make annotations. This is not necessarily the most natural and efficient means for recording ideas given that all the communication in the room relies on oral expression. Taking notes on paper has advantages as it is convenient for most users. Moreover, it is easy to add comments or graphical annotations. On the other hand, the disadvantages of a physical medium such as paper have to do with its limitations for sharing notes and annotations, as well as for copying or archiving sessions.

The emergence of surface computing has led us to envision collaborative environments such as those reviewed in the previous section. Interactive tabletops and multi-touch whiteboards appear as natural options for supporting collaboration and innovation processes, even if not in their entirety. In particular, most of the available applications for multi-tactile surfaces do not effectively support activities that involve adding information in real time or specifying relationships among data in a dynamic manner. Most existing solutions only allow users to view information and perform simple manipulations, such as moving items on display or zooming in and out to increase or decrease the level of detail of what is being presented. Adding or updating information typically needs to be done offline or relies on physical or virtual keyboards which tend to interfere or slow down the innovation process, as participants need to deal with the interface or the physical artifacts rather than only with the ideas being generated and discussed.

**Vozze components**

Vozze integrates three main conceptual components: A voice interface for capturing ideas, a gestural interface for manipulating idea representations, and a collection of workspaces that are created and interlinked as idea generation progresses.

Speech recognition in Vozze makes the process of idea capturing more natural, though it is not intended to replace keyboards or other forms of data entry. In fact, once ideas are captured and represented on the interactive surfaces, users may choose to edit them using a virtual keyboard. Although voice input is a key component that takes advantage of existing developments in speech recognition, what is distinctive in Vozze is its flexible user interface, which allows users to quickly observe how their utterances have been recognized and, if needed, choose the best from several possible interpretations presented in the multi-touch interface.

As for the gestural interface, Vozze allows users to perform rapid actions on objects by taking advantage of multi-touch interfaces. In addition to general gestures for actions such

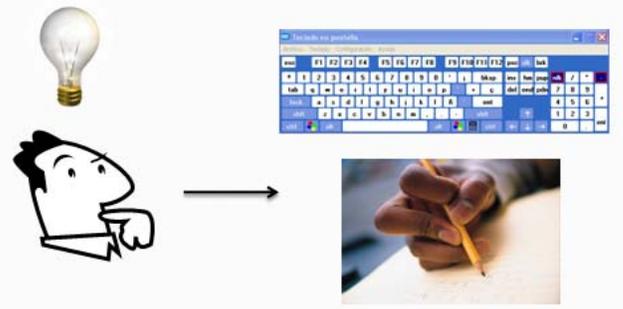
as dragging, rotating, and zooming in and out, Vozze users may also use gestures as shortcuts for actions that occur frequently in innovation activities. These include copying or deleting objects, navigating to and from workspaces, viewing objects' properties, and undoing and redoing actions. It should be noted that using a single-touch surface would significantly limit the number of simple gestures that could be associated to actions (e.g. a triple or a quadruple tap with one finger). We take advantage of multi-touch functionality to rely on a variety of intuitive gestures, as exemplified in the work scenario below.

Workspaces are the third key component of Vozze. They represent our approach to facilitate an enhanced organization of the ideas being generated and recording collaborative sessions. Workspaces are essentially blank areas on multi-tactile surfaces designed to invite users to insert new ideas and to manipulate the ones already on the display. Within workspaces, idea representations can be manipulated by dragging them around, creating multiple copies, resizing them or modifying their appearance (e.g. changing font size or color). By moving objects around an interactive table, users may achieve a better organization of their ideas. Every workspace includes a toolbar that presents controls for both the objects on display and the speech recognition engine.

Additionally, as workspaces are created, they are organized as a network that represents ideas using links that connect related concepts. Links are associated to the text that represents ideas captured via voice, and every link will take the user to a new workspace in which more ideas can be added. Navigation through the existing workspaces is controlled by relying on predefined multi-touch gestures.

**Work scenario**

The development of Vozze is oriented to support scenarios where people work in teams around an interactive surface. Consequently, it responded to the need to provide easy ways to add manipulate and organize the ideas and the knowledge that are generated during a session.



**Figure 1: Conventional innovation scenario.**

Figure 1 illustrates the common scenario in which a user has an idea and needs to type it or write it down in order to

obtain a representation for later reference. The disadvantages of these approaches have been discussed earlier.

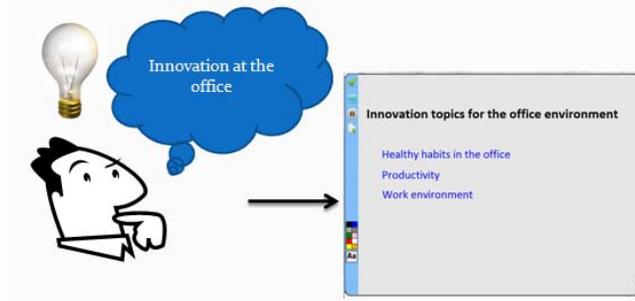


Figure 2: Vozze scenario

One important contribution of Vozze is that voice is its main way to capture ideas. If a group member has an idea, he or she just needs to talk, and the speech recognition component will capture the idea and present it as text on the multi-touch display. In this way, the user will not have to worry about typing or dictating the idea to a partner. This is as easy as talking and seeing ideas represented on the screen, and manipulating them within a workspace.

Figure 2 shows how Vozze works: A user has an idea, speaks up his mind, sees textual representations on the screen and may start working on it by arranging text items spatially, and elaborating on the idea by linking the text to newly created workspaces. Users may also modify the format, size, and position of any items on the workspace.

Figure 3 shows a scenario in which workspaces are created and inter-linked. In our example, users around an interactive surface are generating ideas about innovation in the office environment. Upon starting the session, a blank workspace is presented and participants have used the speech recognition component to enter a title (“Innovation topics for the office environment”) and three main sub-topics: healthy habits in the office, productivity, and work environment. A participant decides to elaborate on the first topic, so she uses a gesture (double tap with two fingers) on the corresponding text item to create a new workspace that is linked to “Healthy habits in the office.” By adding new workspaces and links between them, users can organize ideas in a way that closely resembles the way in which they are generated during the session.

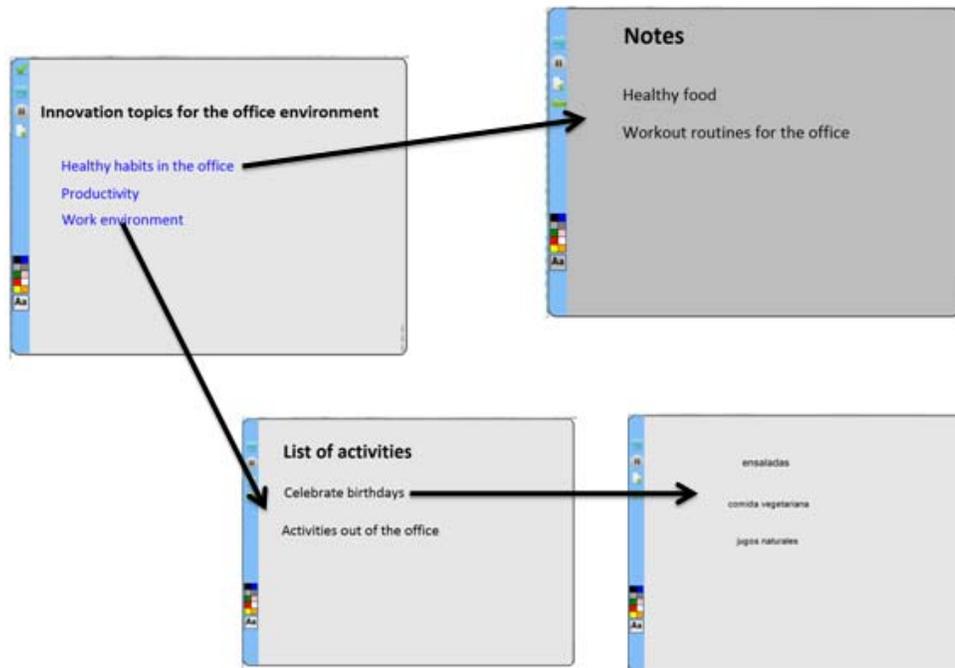


Figure 3: Links and workspaces.

In the new workspace, shown on the top right area of Figure 3, session participants have added a title (*Notes*) and two issues they are discussing about the topic at hand (*Healthy food and Workout routines for the office*). Users

also can use simple gestures in order to navigate between workspaces, so they may easily move back and forth between discussion topics and subtopics.

Figure 3 also illustrates, at the bottom, a two-level elaboration of the “*Work environment*” item mentioned in the initial workspace. In this case, a list of activities has been created and linked to the third item in the first workspace and an additional workspace has been created and linked to the first item of the list of activities. As noted in the figure, users may continue to add items (by just speaking up), create, link and populate new workspaces, or navigate around this representation of their ideas (by using predefined gestures).

Only one workspace is displayed at any given time. In this sense, we have followed the style introduced by KMS, which aims to reduce disorientation by emphasizing speed for displaying existing or newly created workspaces.

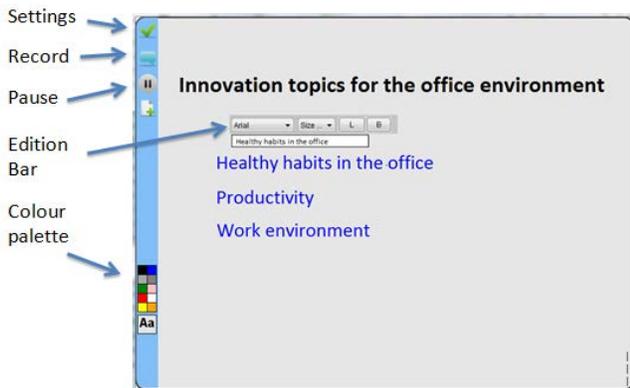


Figure 4: Vozze’s graphical interface.

**VOZZE INTERFACE**

Vozze’s graphical interface, illustrated in Figure 4, is quite simple. It includes a toolbar on the left side which provides options to start (*Record*) or stop (*Pause*) the speech recognition engine, and a color palette that can be used to change the color of items or the workspace’s background.

The icon on the top left corner can be selected to change the settings of the interface, whereas the bottommost icon of this group, a blank page with a plus sign, provides an alternative for adding text items by typing if users prefer not to use voice. The figure also shows the formatting options that are presented when the appropriate gesture is applied to an object on display. Options for text items include changing the font family, size or emphasis, as well as an alternative to the gestural interface to create newly linked workspaces.

Current speech recognition techniques work really well most of the time, particularly after some training. However, some inaccuracies still occur in most implementations. In order to deal with these recognition errors, we decided to include mechanisms that facilitate error correction while

users are talking. A first method to correct errors is provided by the speech recognition engine, which will provide three different results for each recognized phrase. These options are displayed on the bottom left corner of Vozze’s interface, so the user may pick the correct option and then manipulate the corresponding text within the workspace, as illustrated in Figure 5.

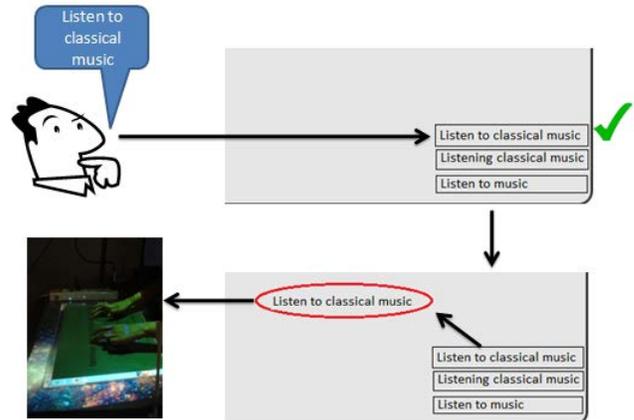


Figure 5: Vozze Functionality

In addition to the three different results provided by the speech recognition engine, Vozze also allows users to edit text items by using a regular or virtual keyboard. This option is helpful when minor errors are introduced by the speech recognition engine and editing may be faster than re-capturing an entire idea via voice. This way, one of the results produced by the speech recognition engine can be selected (the one closest to what the user intended to say), and do minor editing instead of typing the entire idea.

Figure 5 summarizes Vozze’s operation: A session participant has an idea and says it out loud. Then the speech recognition engine will recognize the words and will display three possible phrases on the bottom left corner of the display. The user touches the option that matches what was said, or edits the closest match. After that, the user drags the text item to a convenient area within the workspace.

**PROTOTYPICAL IMPLEMENTATION**

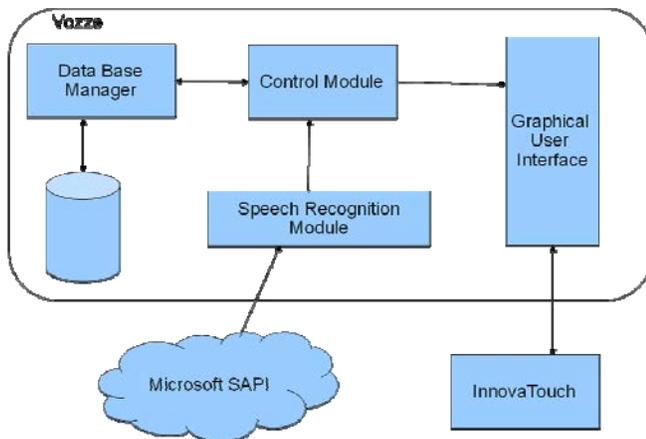
We have implemented a proof-of-concept version of Vozze by integrating some existing technologies and developing some needed components. The resulting setup is shown in Figure 6. In addition to a microphone, we have used an interactive table that allows for a small number of (2 to 4) collaborators to participate in an innovation session. Although several users can participate in a work session, our current implementation does not associate contributions to specific individuals.



**Figure 6: A proof-of-concept implementation of Vozze.**

Figure 7 portrays the main components comprised by Vozze: A graphical user interface, a speech recognition module, a database manager and a control module.

The graphical user interface was developed using JavaFX and implements the features and controls described in the previous section. Every workspace is an object for JavaFX and text items become JavaFX nodes. Workspaces share a thread with the speech recognition module, so results are assigned to the appropriate workspace.



**Figure 7: The architecture of Vozze.**

The database manager provides persistence functionality to store and retrieve workspaces and their associated data, which include appearance metadata as well as the items and links created by users during each session. The MySQL relational database manages has been used to provide data persistence during work sessions.

The speech recognition module is in charge of collecting the results of the speech recognition engine and also of providing basic functions such as starting, pausing and stopping the speech recognition engine. Moreover, this

module obtains the text triplets that result from the speech recognition engine, which are presented to the user for selection and edition (if needed).

The control module manages the entire interface by associating the intended actions to multi-touch gestures, obtaining results from the speech recognition engine, organizing the text on the display and managing links to other workspaces.

In order to provide multi-touch gesture recognition, Vozze relies on InnovaTouch, a simple framework currently under development in our lab for supporting operations on multi-tactile surfaces. InnovaTouch proposes a set of gestures for common actions throughout applications as well as gestures that are more specific for the innovation scenarios which we are aiming to support. InnovaTouch is described in detail in [5].

Speech recognition is implemented via the TalkingJava libraries that provide an API for Java to obtain the results of the speech recognition engine installed in the computer. In the case of Vozze, we used the speech recognition engine by Windows 7, which is referred to as Microsoft Speech Recognition API (SAPI).

Ultimately, Vozze integrated these concepts and presents them to the user as a natural means to support early stages of an innovation process using multi-touch surfaces. Although the interactive table we are using does not impose limitations on the number of contact points that can be enabled, Vozze’s underlying InnovaTouch gesture service was designed to work with interactive surfaces with up to 4 contact points. This makes it possible for applications such as Vozze to work with more limited surfaces, such as those implemented by using the Nintendo Wiimote<sup>1</sup>.

**EVALUATION**

In order to assess the potential, applicability and limitations of Vozze, we conducted a study with actual users that performed various tasks. The study focused on evaluating and obtaining feedback from users on the following;

- The general usability of Vozze
- The potential of voice as a method for capturing new ideas when interactive surfaces are used
- The advantages and cognitive overload, if any, implied by the use of dynamically inter-linked workspaces for organizing ideas.
- The areas of opportunity to improve Vozze and to enhance collaboration and innovation around shared interactive surfaces.

<sup>1</sup> Projects using the Wiimote, by Johnny Lee. <http://johnnylee.net/projects/wii>. Last accessed on June 10<sup>th</sup>, 2011.

The seven subjects who participated in the study came from areas as diverse as Actuarial Science, Communication, Information Design, Information Systems and Computer Science, and Mechatronics. There were 3 females and 4 males, all between ages 18 and 30. All sessions were recorded on video and also using a screencast tool.

The evaluation setup is depicted in Figure 8. Given the prototypical nature of both our hardware and software components, we still had to maintain a very controlled environment with respect to lighting conditions, and we had to find a combination for background and foreground colors that helped produce a reasonable gesture recognition rate. Still, users were able to perform the tasks requested and speech recognition worked reasonably well.



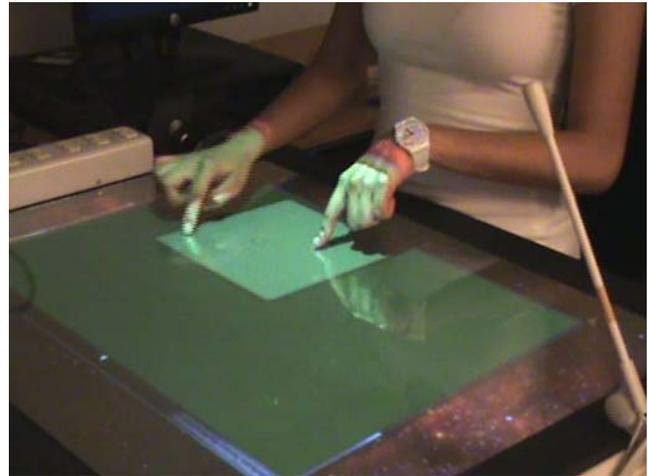
**Figure 8: Vozze Interface**

During an initial part of the study, users went through a stage of adaptation and learning of the set of gestures defined for Vozze, and also with the responsiveness of the interactive table. In this stage, subjects were presented with a simple application designed to test InnovaTouch (see Figure 9) and were asked to try gestures for copying, zooming, rotating, and modifying object properties, among others.

After the adaptation stage, participants worked on predefined tasks and, lastly, they had some time to try and explore the prototype without any restrictions.

The tasks defined for the usability study assumed a scenario in which Vozze supported an innovation session. In this session participants would try to brainstorm about a predefined topic, in our case, “The Healthy Office”. In one of the tasks, for example, subjects were asked to suggest two ways in which activities performed in an office setting could be made more aware of the environment. Later on, one of these ideas had to be further elaborated. In this process, subjects were required to express their ideas to experiment with speech recognition and correction,

manipulate text items to observe the use of gestures, and various editing and formatting tasks, as well as arranging items within a workspace. Elaborating on ideas required creating and linking new workspaces.



**Figure 9: Exploring gestures used in Vozze.**

## RESULTS

We now discuss the results observed during the study, starting with a summary of responses to post-test questionnaires, which is summarized in Figure 10. Numerical responses ranged from 1 to 4 (totally disagree to totally agree). The responses were added up and then normalized to a 0-100 scale.

As observed in Figure 10, our approach to using voice was very welcome by participants. All of the subjects totally agree (hence the 100 rating) that it is helpful to use speech as a method to capture new ideas on interactive surfaces. In the post-test questionnaire, one of the subjects noted: “It would take much longer to enter the entire text. [Even when not completely accurate] it is much easier to use your fingers and select the recognized text to correct it.” Another subject added: “This saves typing time!”, whereas other comments noted that “It will be helpful in collaborative work”, “It will make work more agile”, “Precision [in voice recognition] and speed are nice”, and “It is nice not having to pause to take notes.”

The notion of inter-linked workspaces also is promising, as the method for organizing ideas was given an 82 rating. Comments by subjects indicated that workspaces are simple and practical to get a clear idea of the structure of the topics discussed during a collaborative session. The Vozze concept as a solution to support innovation process was given an 85 rating, which also is a positive result. In this regard, comments from users included: “It is nice to go into detail without having to open new files”, “This helps to organize your ideas”, “the organization is very intuitive”, and “It is a good way to keep a record of the group’s discussion.”

When comparing keyboards to voice as methods for capturing ideas, our subjects preferred the latter (85 rating), and they commented that voice is more natural and less intrusive when you are using interactive surfaces, whereas a virtual keyboard seems cumbersome. Even though speech recognition is far from perfect, the engine used in Vozze was good enough to demonstrate the advantages of using voice in these scenarios, and users also rated it highly (89).

It is fair to mention that subjects were not totally pleased with the current state of editing functionality, as the correction options for speech recognition received an aggregated rating of 67. Formatting options also are an area that requires a lot of improvement, as users have great expectations on how they can present text items and showed a low acceptance of currently available functionality (about 42). It is clear then that additional ways to correct the results of speech recognition need to be explored. Also, although presentation is not central to our concept, we are aware that formatting options impact how users present their ideas and how they perceive Vozze.

Thus, it is necessary to work on improving the format toolbar.

Other responses to open questions show user enthusiasm regarding the potential of speech recognition, and even suggested ways to deal with environmental noise. For example, some participants suggested including some graphical indication, icons perhaps, that show the activity of the speech recognizer and provide feedback about its progress.

Some of the predefined gestures took more time than expected to be learned, and at times users had to rely on cue cards we provided to remember the appropriate gestures. We have defined this as an area of opportunity for our research, as we are interested in determining an optimal combination of gestures and graphical interface elements that would promote a natural and agile interaction. We did observe, though, that most users were able to master the proposed gesture set, and found it practical to navigate through the workspaces, create links and manipulate objects as needed.

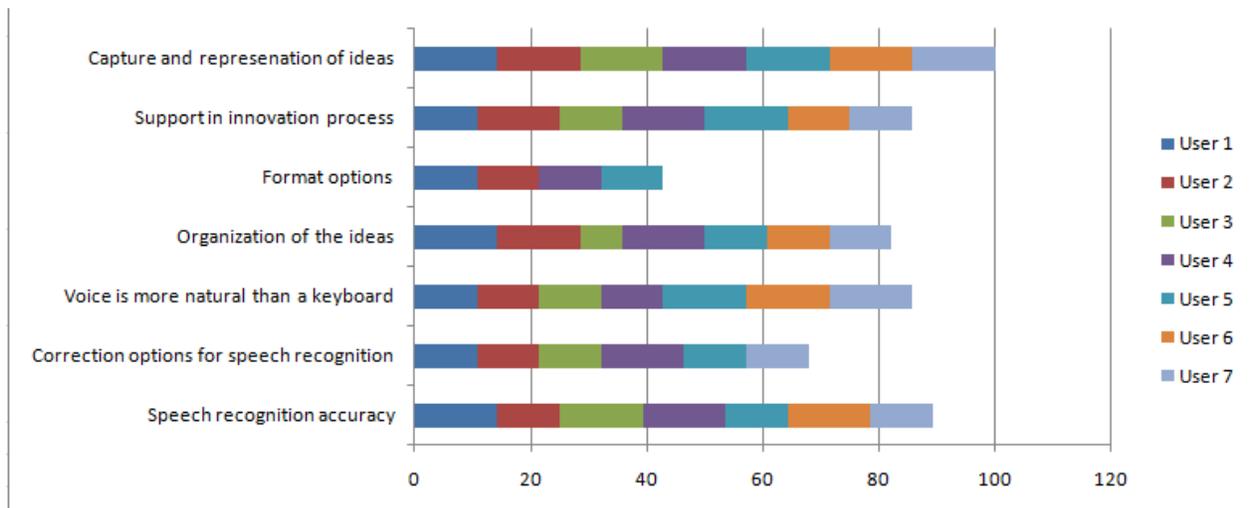


Figure 10: Results from studying the use of Vozze.

**CONCLUSIONS AND FUTURE WORK**

Creative group activities that are part of innovation processes have relied on conventional artifacts such as paper flipcharts, markers and post-it notes in order to reify and facilitate the discussion and generation of ideas. While these are intuitive tools, they also present disadvantages for preserving what is discussed, for promoting participation from all members in a group, and for facilitating the transition to other stages of the innovation process. It has not been until recently that natural user interfaces are presenting opportunities for supporting such creative activities with information technologies. Interactive surfaces and embodied interfaces are among the technologies that show particular potential to be incorporated into innovation activities. Work still is needed

to make this incorporation smooth and natural. We have developed Vozze, an environment that makes a contribution in that direction by exploring the use of speech recognition as a means for capturing ideas, providing textual representations on interactive surfaces, and facilitating the manipulation and association of ideas. We have built a prototypical implementation of Vozze and our observations of users in a controlled environment are encouraging regarding the potential of our approach to a natural interaction scenario.

From studying users and obtaining direct feedback from them, and from our interaction with innovation teams, we have derived a number of areas for further work and research. Clearly, it is necessary to focus on the incorporation of improved speech recognition components,

which may take advantage of custom dictionaries, voice user profiles and emerging algorithms. The more direct becomes the transition between the user's idea and its representation on the display, the more natural the interface will be.

Since speech recognition is likely to remain imperfect for some time, we need to make sure users have enhanced correction mechanisms at their disposal. Predictive text is one direction we can explore for text correction using virtual or physical keyboards. Simple context awareness has shown to improve speech recognition using prediction models. User profiles also would allow for identifying specific users and, therefore, allow Vozze to keep track of who is participating in each session.

As noted earlier, users demand more options to format text and personalize their environment. In general, we need to study visual aspects of the interface in order to improve its support for user needs and hence its user acceptability.

An ongoing project in our lab already is investigating the use of several forms of menus (e.g. fan and half-pie menus) that will pop up at any moment when certain gestures are applied on the multi-tactile screen. We aim to support various data entry modalities, including speech, gestures and graphical interfaces. We expect users will determine what is optimal (more natural, more efficient or more effective) according to their individual needs and preferences. In this new project we also are including functionality for adding images and free-hand drawing, which also are much needed actions in innovation settings.

We believe voice and gesture interfaces are a necessary complement for graphical user interfaces and conventional input devices towards the goal of providing more natural and intuitive technological support for all sorts of activities, including creative and innovative group collaboration. The results from our work on Vozze provide evidence of this potential confluence and motivate further research.

#### ACKNOWLEDGMENTS

This project has been supported by the CUDI-Conacyt Program for Advanced Applications and the Conacyt Mexico-Germany (DLR) Bilateral Cooperation Program. We also gratefully acknowledge support for this work from the French-Mexican Laboratory of Computer Science and Automatic Control (LAFMIA-CNRS UMI 3175). Multi-touch equipment and tracking software has been provided by ITAI and EDIS (edis.mx) through a cooperation agreement with Universidad de las Américas Puebla and the ICT Lab (ict.udlap.mx).

#### REFERENCES

1. Akscyn, R. M., McCracken, D. L., and Yoder, E. A. 1988. KMS: a distributed hypermedia system for managing knowledge in organizations. *Commun. ACM* 31, 7 (Jul. 1988), 820-835.
2. Castellina, E., Corno, F. y Pellegrino, P. 2008. Integrated speech and gaze control for realistic desktop environments. *Proceedings of the 2008 symposium on Eye tracking research & applications* (Savannah, Georgia), 79-82.
3. Cohen, P., Coulston, R. and Krout, K. 2002. Multimodal Interaction During Multiparty Dialogues: Initial Results. *Proceedings of the 4th IEEE international Conference on Multimodal interfaces - Volume 00* (October, 2002), 448.
4. Dragon Naturally Speaking. <http://www.scansoft.com/naturallyspeaking>. Last accessed on June 9<sup>th</sup>, 2011.
5. Fuentes, S. 2011. InnovaTouch: Servicio de ademanos multi-táctiles para actividades de innovación. B.Eng. Thesis. Dept. of Computing, Electronics and Mechatronics. Universidad de las Américas Puebla (Dec.).
6. Haller, M., Leitner, J., Seifried, T., Wallace, J., Scott, S., Ritcher, C., Brandl, P., Gokcezade, A. y Hunter, S. 2010. The NiCE Discussion Room: Integrating Paper and Digital Media to Support Co-Located Group Meetings. *Proceedings of the 28th international conference on human factors in computing systems* (Atlanta, USA), 609-618.
7. Holzapfel, H., Nickel, K. y Stiefelhagen, R. 2004. Implementation and evaluation of a constraint-based multimodal fusion system for speech and 3D pointing gestures. *Proceedings of the 6th international conference on Multimodal interfaces* (State College, PA, USA), 175-182.
8. Liu, P. y Soong, F. 2006. Word graph based speech recognition error correction by handwriting input. *Proceedings of the 8th international conference on Multimodal interfaces* (Banff, Alberta, Canada), 339-346.
9. Loton T., *Talking Java! Add speech capability to your Java 1.3 applications and applets*. <http://www.javaworld.com/javaworld/jw-08-2001/jw-0817-javataalk.html>. Last accessed on June 9<sup>th</sup>, 2011.
10. Microsoft Speech SDK 5.1. <http://www.microsoft.com/speech/download/sdk51>. Last accessed on June 9<sup>th</sup>, 2011.
11. NASA Sub-Vocal Speech Research. <http://www.nasa.gov/centers/ames/news/releases/2004/subvocal/subvocal.html>. Last accessed on June 9<sup>th</sup>, 2011.
12. Rogers, Yvonne; Lim, Youn-Kyung and Hazlewood, William (2006). Extending tabletops to support flexible collaborative interactions. *In: Proceedings of the First IEEE International Workshop on Horizontal Interactive Human-Computer Systems*, 5-7 January 2006, Adelaide, Australia.

13. Sánchez, J. A., Strazzulla, D., and Paredes, R. 2008. Enhancing interaction and collaboration in multimedia rooms with multilayered annotations and telepointers. *Proceedings of the VIII Brazilian Symposium on Human Factors in Computing Systems* (Porto Alegre, Brazil, October), 117-125.
14. Speech Software at CMU. <http://www.speech.cs.cmu.edu/hephaestus.html>. Last accessed on June 9<sup>th</sup>, 2011.
15. Streeter T., Open Source Speech Interaction with the Voce Library. Virtual Reality Applications Center, Iowa State University, Ames, IA. <http://voce.sourceforge.net/files/VoceWhitePaper.pdf>. Last accessed on June 9<sup>th</sup>, 2011.
16. The VoiceCode Project. <http://www.nrc-cnrc.gc.ca/eng/projects/iit/voce-programming.html>. Last accessed on June 9<sup>th</sup>, 2011.
17. Tse, E., Greenberg, S. y Shen, C. 2006. GSI demo: multiuser gesture/speech interaction over digital tables by wrapping single user applications. *Proceedings of the 8th international conference on Multimodal interfaces* (Alberta, Canada), 76-83.
18. Tse, E., Shen, C., Greenberg, S. y Forlines, C. 2008. Enabling interaction with single user applications through speech and gestures on a multi-user tabletop. *Proceedings of the working conference on Advanced visual interfaces* (Venezia, Italy), 336-343.
19. Vertanen, K, y MacKay, D. 2010. Speech dasher: fast writing using speech and gaze. *Proceedings of the 28th international conference on Human factors in computing systems* (Atlanta, USA), 595-598.
20. Wheeland A. How to Use Voice Chat on WOW. [http://www.ehow.com/how\\_7415210\\_use-voice-chat-wow.html](http://www.ehow.com/how_7415210_use-voice-chat-wow.html). Last accessed on June 9<sup>th</sup>, 2011.