

## Case Study: Providing Enhanced Content Accessibility to Images for Visually Impaired Users

Authors: Thimoty Barbieri, Licia Sbattella, [barbieri@elet.polimi.it](mailto:barbieri@elet.polimi.it), [sbattell@elet.polimi.it](mailto:sbattell@elet.polimi.it)

### **Foreword and Motivations**

Several web sites, such as e-learning platforms, online museums, etc. are available via the web, and allow users and visitors to learn at a distance. Several e-learning platforms or in general informative web sites (such as museum web sites) have also started to make their web site “accessible”, making them more navigable for (visually) impaired users. However, accessibility in the current conception concerns navigation within the site, but does not help to effectively peruse the content in the site itself. For specific learning contexts, such as geography, math, history, or in museum web sites, access to images and graphics that carry significant meanings which are hardly translatable into words is barren to visually impaired users. In this case, providing a textual alternative like it is suggested in the accessibility guidelines may not prove sufficient, let alone satisfactory for the user. A solution to this problem is offering the image information in a multimodal way, enriching the graphical information with haptic (tactile) feedback and with aural (audio) cues. In a multimodal version of an image, it could be possible to allow feature identification with a correlation to vibrational patterns, with corresponding voice annotations, so when for example the user hovers on a city in a map, the corresponding tactile cue is felt, and a voice reads the name of the city.

In using haptics feedback, our study focused in trying to provide to the visually impaired a mean to create a mental map of the spatial organization of visual elements, with two different purposes:

- Provide navigational clues, giving information on how an object is shaped, or how a seeing user perceives it, in terms of main area of content, contours, dimensions, etc.
- Provide tactile cues to describe the content of an image, allowing interactive tactile exploration of its main features (namely, the features that the image wants to convey to the reader, for example the cities in a map, or the mutual position of characters or objects in a painting).

A first attempt was made using the Immersion TouchWare™ technology, and with off-the-shelf iFeel™ mice, or force feedback joysticks, such as Microsoft's. However, for blind users the use of the mouse or of joysticks is practically unknown, and, besides sharing the initial difficulties in using a mouse typical of all seeing novice users, the impossibility to relate the cursor speed and acceleration on screen to the movement of the hand prevents the blind user from getting useful information.

Exploration of the features of an image (like a geographical map or a painting) in this way proves to be impossible: a disadvantage in the use of the mouse, is that it is difficult to reproduce movements and assess the distance of the movement without seeing it, coupling the action on the mouse with the distance on the screen. Therefore, it requires a long, tiresome and extensive training to discover where the cities are and reproduce paths across them blindly. Also experiments using a joystick have been carried, but did not prove very successful. The joystick allows to quantize movement steps and understand the exact direction in which the movement has been expressed, however the interaction proves counter-intuitive and creating a mental spatial map of the image using step-by-step movements with a joystick is extremely difficult.

For this reason, it was necessary to study a different way to let the blind user interact with web-based content in order to allow some sort of spatial exploration of images.

### **A special device for image exploring**

During this study, we created a device which could allow multimodal exploration of web images in a much more direct way. This device, called “AudioTact”, uses the two stereo audio channels to offer a combination of audio and tactile stimula during exploration with a touch screen of an image downloaded via the web. The Web Image is pre-prepared in a way that it can generate sounds which are produced on the left channel by regular audio headphones and on the right channel by a haptic transducer which has to be worn on the finger or attached to the pen of a graphical tablet. Following a given exploration method, the visually impaired user can autonomously extract information from the features of the image.

The “AudioTact” device and related methodology allow to download from a web page a raster image (e.g. JPG or PNG), enriched with linework shapes (Image Maps), created by way of the IMAGE MAP tag in HTML. Using some crafted JavaScript callbacks, given areas of the Image Map are linked to a .wav stereo sound specifically prepared, which carries in the right channel a synthesised voice which describes vocally the area linked to the image, and in the left channel a waveform with high intensity and low frequency (usually a sine, in which amplitude is varied to create different intensities of tactile perception). The user starts exploring the image using a regular PC with a device which

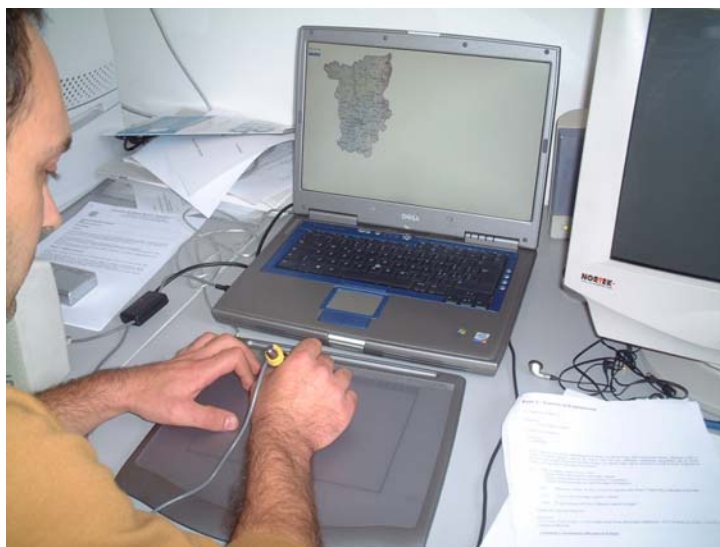
allows direct and absolute pointing on the screen area, for example using a touch screen with a finger; a pen on a Tablet PC; a pen on a drawing tablet. On the finger or on the pen, that is on the device which is used for absolute pointing, is mounted the vibrational transducer of the AudioTact device. This AudioTact device is connected to the audio outlet for the PC headphones, and it is fitted with an adapter which allows to connect regular PC headphones, and of a vibrational cell which is to be worn on the finger or attached to the tablet's pen. In this way, the image being prepared for tactile exploration and downloaded from the web, the user using the AudioTact can explore the image itself dragging over it the finger or the pen, meanwhile receiving audio cues from the headphones and tactile cues from the vibrational cell. Combining the two cues, the user is able to move within the image and understand its features.

The best exploration strategy is the following: the left hand is kept open and resting over the surface of the screen, in a way that the borders of the screen itself can be felt. The right hand wears the AudioTact device and is used for exploration. The right hand is moved in a way that it always keeps contact with the left hand, to preserve an absolute reference of the screen's real estate. The hands move on the screen's surface at a speed of approx. 5 cm / sec. The user starts moving along the diagonals of the area being explored to determine the bounding region in which information is present (detectable by the presence or absence of tactile cues). Once the general bounding area of the image is understood, the user starts exploring from a corner where the tactile cues reveals the presence of the border of a feature of an image, and starts following this border performing small zig-zag movements, until a change of direction in the border is detected. At this point the users look in the immediate surroundings with small circle 360 degrees movements, until the continuation of the border is found. The path is followed several times in this way, until its structure is understood, and the perception of the shape is consolidated by following it several more times with straight movements. At this point the shape of the pictures is understood. The whole process requires from 180 to 300 seconds for shapes of average complexity to a non-trained user. Image exploration of images for which it is not required shape exploration but exploration of features dispersed within an area (geographical maps, mathematical functions), is proportional to the number of informational elements with which the user wants to build its mental map of the spatial layout of these elements.

There are other implementations for sending force feedback stimulations over the web to users (implemented by Immersion corp.). These systems, however, require the installation of a client software on the PC of the user, while the vibrational cell is inserted into a mouse or a joystick. This approach defeats completely the use of the haptic cues for exploring images, because it decouples the cue from the notion of the absolute position of the user on the screen. The unsatisfactory experiments using this system led to the development of our device and method. The innovative issue in our approach is that we used audio output separated in the two channels of a stereo signal to provide concurrent audio and tactile cues, the tactile cues being created by a mechanical vibrational cell controlled by an audio signal. In this way it is simple to transmit over the web the necessary signals without adding other software at the user's side. Moreover, by moving the vibrational cell from a relative pointing device directly on the user or on the pen, we provide a completely different concept on how to use vibrational cues, not as "force feedback" output, but as a mean to drive autonomous exploration of images. It is important to know that the combined use of aural and vibrational cues is crucial for a successful exploration of the image. In fact, using only the aural cue is not sufficient to determine the exact position on the plane of a specific point of a contour during blind exploration. The pronunciation of the name or of the sentence explaining the specific spot reached by the user, takes a certain amount of time to be uttered. During this time, the blind explorer moves the pen and therefore is not able to determine the exact spot in which the specific object lies, within the range of the movement performed in the meantime. By using the vibration, which occurs only on the specific spot, the blind user is able to pinpoint the exact position and to add it as an information for building his/her mental map.

Another main advantage of this approach, is that images can be downloaded via the web and explored autonomously. A blind user trained in this technique could visit learning site rich with graphic contents (such as sites about geography, or art sites), and extract relevant information from the graphics. This would be an advantage over the current technique, which is to realize an expensive cardboard relief version of the image. The web image, being virtual, does not require production and material costs, and can be quickly downloaded at no cost. Moreover, it can be dynamic, meaning that its features could change to allow the exploration of different "levels" of elements.

The main disadvantages are related to the fact that most visually impaired users are educated in using *relief*. Exploring relief is a tactile sensation, while our device gives *vibrational* outputs. Therefore, the process of making a spatial map of the image is completely different and requires a higher level of attention. Users with difficulties in concentrating during the exploration phase have problems in relating the virtual experience of our device with the physical feedback provided by the tactile exploration of a relief. To this end, we found that the exploration could be made easier by adding a teflon sheet when using a pen with a graphical tablet. The teflon sheet is such that it can be engraved with the pen, and the marks made on it can be felt in relief. In this way, users can engrave the teflon sheet when, during the virtual exploration, they feel a vibration from the pen and a vocal suggestions from the audio speakers, and subsequently touch the relief they had made to help them creating their mental spatial map.



Figures 1-3: Experimentation of the AudioTact device. *Top*: a picture of a pen with the vibro-audio device attached. *Middle*: a visually impaired user explores the virtual image using the left hand for spatial references and the right hand for scanning for aural and vibrational cues. *Bottom*: Exploration of a geographical map.

## Experimentation and Results

This approach was tested in a case study which involved ten blind students of the Institute of the Blind in Milan. The test provided a training phase, in which the student could get confident with the method, and a second part in which it was required to use the system to extract new information from unknown images.

During the training stage, we explained to each student how the method worked, and we announced that he/she was going to explore a simple shape: a square. We then let the student get acquainted with the pen, the table and the vibrational feedback until he/she was able to re-trace correctly the square shape on the table.

In a second step of the training, we changed the image to a map of the Region of Lombardy, that is the Region in which the students live and study. We let the students exploring the map, learning how to use the vibrational cues together with the aural cues. This step was important to explain to the student the importance of having the aural and vibro cues coupled together. At first, the students were using only the aural cues to explore the image (coupling their notion of position of an object with the synthesised voice describing it). Soon enough, they realized that this was not correct, because of the time needed to utter the name and of the space covered by their movement of exploration in the meantime, and therefore learned advantageously to use the vibrational cue to fix the spot while using the aural cue to determine its nature.

During the exploration of the Map of Lombardy we asked three questions: a) Locate the City of Milan; b) Locate the Po River; c) Locate the Garda Lake, and we took note of the total timing needed for exploration and the number of correct answers (Table 1). At the end of the training session, we required the students to extract autonomously information from two images, a simple shape of a triangle and another geographical map, representing a region of which they did not have any previous knowledge (the map of the Region of Perm, in the Russian Federation).

We first loaded the image of the triangle, and we asked the students to try to understand the shape by following its contours, and we took note of the time needed for the students to explore the shape and recognize it (or forfeit, declaring that they couldn't understand). Subsequently, we explained that they were going to learn the geography of a region of the Russian Federation, and read in advance the questions that they should have answered:

- a) Does the Kama River flow in the center of the region?
- b) Does the Kama River flow across the city of Perm'?
- c) Does the Kama River flow across the city of Kungur?
- d) Place in order, from South to North, the following cities: Perm', Chaikovskij, Solikamsk, Berezniki
- e) Which city is farthest from Perm', Solikamsk or Kungur?

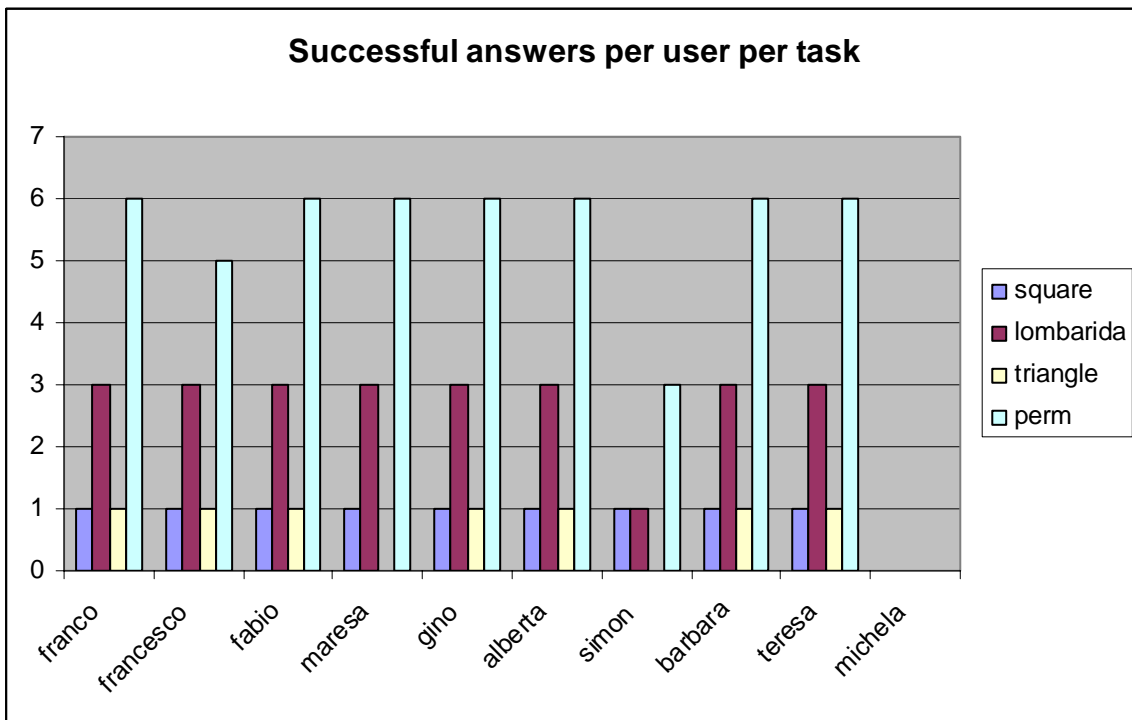
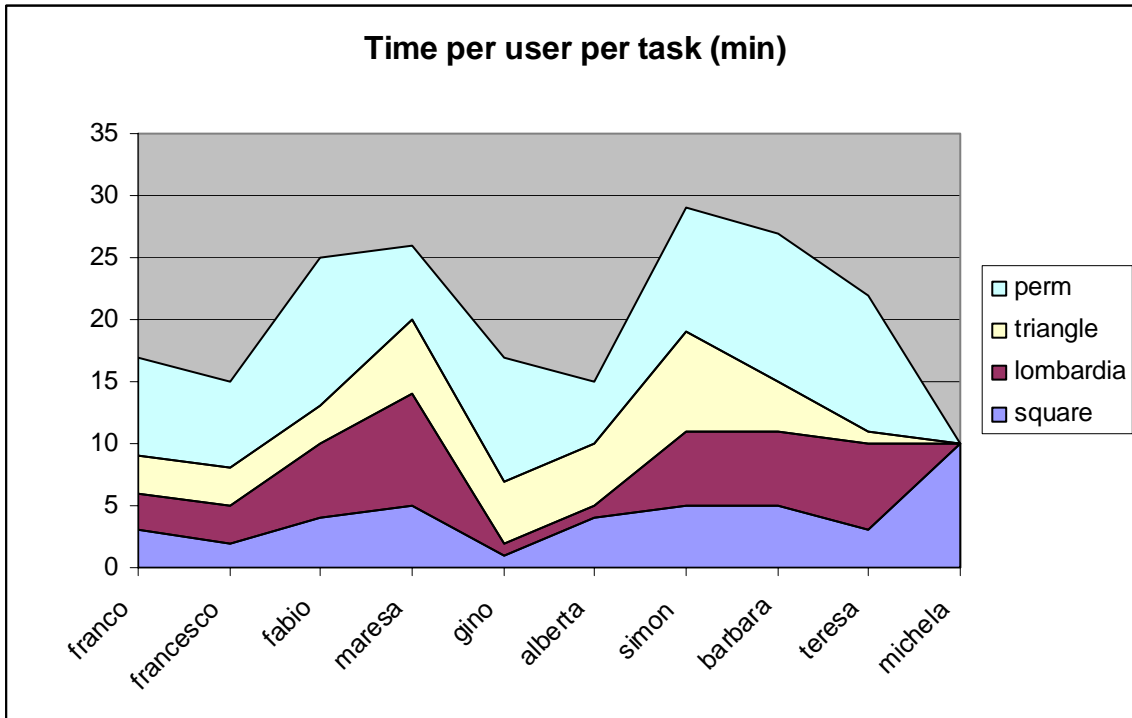
After some exploration, the students started to answer the questions, and we noted the time and the number of correct answers they gave. At the end of the experiment, we asked the users if they thought they could use such a method for a web-based course of Geometry or of Geography.

The results are collected in Table 1. The table shows that all users except one easily recognized the square (in about 4 minutes). One user declared that she did not like the tablet being smooth and that she only could explore using relief. For this reason we are going to add the teflon sheet, engravable with the pen, in the next experiments, as explained before. Autonomous exploration gave comparable results (most of the students answered correctly after some time of exploration), meaning that they could extract fresh information from the image using this method.

Shape exploration proved to be more difficult than map exploration, and for this reason most of the students indicated image exploration (map, pictures) as a better application for the method, than contour recognition.

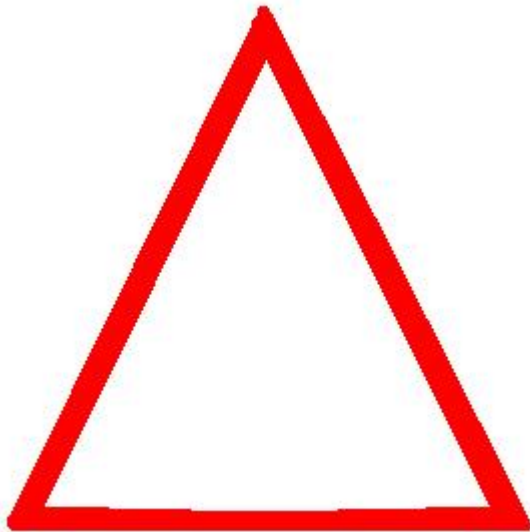
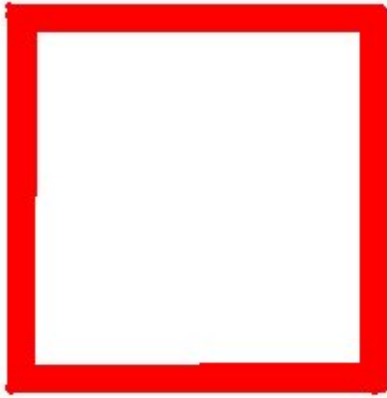
<b>Test of Usage over 10 blind users</b>				
	<b>% Of users</b>	<b>Average Time (min)</b>	<b>Max Time (min)</b>	<b>Min Time (min)</b>
Recognizes Square	90,00%	4,2	10	1
Avg % Correct Answers on Lombardia Region	83,33%	4,7	9	1
Recognizes Triangle	70,00%	4,2	8	1
Avg % Correct Answers on Perm Region	81,25%	9,0	12	5
Want to use for Geometry	60,00%			
Want to use for Geography	80,00%			

Table 1. Experiment of blind exploration of geometrical shapes and geographical maps. The square shape and the map of Lombardia were explored during the training session, assisted by a trainer. The triangle shape and the map of Perm Region were explored autonomously and without previous knowledge of them. The table presents the % of users over the total that had a success in the specific task, and the average, maximum and minimum time overall needed by the users to accomplish the tasks.



Figures 4-5: The graphs show the overall time for each student to complete the whole test, with the partial time needed for each single task. For each task, is shown the number of correct answers by each student. For the square and the triangle, a value of 1 means that the shape was successfully recognized. For the task of the exploration of the Map of Lombardia, there were three questions to be answered. For the task of exploration of the Map of Perm' Region, there were six questions to be answered. The graphs show how all students except one completed all the tasks successfully and by mostly answering correctly to all the questions.





Figures 6-9: The images used during the experimentation. From top to bottom and from left to right: the square and the map of Regione Lombardia, used during the training phase. The triangle and the map of the Region of Perm used during the second stage, in which blind students were required to extract autonomously information from the images downloaded via the web.

## Conclusions and Future Work

The method looks promising and it was welcomed by most of the students. Most of the students laughed with elation when they discovered how they could tell by themselves the distances and the related positions between cities, without anybody helping them. It is, however, a method which requires good concentration abilities during the process of exploration in order to be able to form a correct mental map of the image. Some students did not have the ability or the will to concentrate for the necessary time to extract the information, and complained that they needed the relief. However, students that were able to concentrate on the tasks, completed the exploration in a few minutes and were able to give correct answers to (rather difficult) questions. It is however true, that the time needed for vibro-aural exploration is several times longer than the one needed with relief exploration. However, relief exploration with cardboards does not provide interactive aural cues, and always requires a) possibility to cover production and material cost of the cardboard; b) availability of a tutor describing the features which are currently touched on the relief, and therefore lacks the possibility to extract autonomously information which can be potentially created in real-time and distributed remotely via the web.

To improve the transition between the virtual stimuli provided by vibro-aural exploration and the physical stimuli of relief exploration we propose to use the pen and the tablet in combination with a teflon sheet which can be engraved with the pen during the exploration. In this way, users can engrave the sheet to draw the information they discover via the vibro-aural exploration, being able to create their own relief from information interactively downloaded from the web, and, more important, being able to explore the relief autonomously after they have created it, because of the aural and vibro cues provided via the web.

Other improvements will be on the possibility to add / remove quantity of information from the image, or control the nature of information contained by the image, for example during exploration of complex maps. Another domain that seems interesting to users we have been discussing with, is the possibility to understand the main spatial features of images of buildings and of famous paintings.

An example on how vibro-aural exploration can be used in an art web site to make more accessible graphical content to visually impaired users is in the following example (fig. 10): in this example, the Last Supper is contained in a web page of an art museum explaining how the masterpiece is composed and the significance of every visual element in it, a token of knowledge shared by many (especially in the trail of the last "Da Vinci Code" bestsellers), from which sensorially impaired users are left out. This particular version of the image contains also vibrational cues laid out in contours and paths, to highlight the most important features of the picture. These cues are complemented with audio cues (pre-recorded or generated with a high-quality, multilanguage Text To Speech engine). The user downloads the cue-augmented image from the accessible museum site, uses the AudioTact device (wearing the vibro transducer on the finger or on his/her pen), and starts exploring. He/she can then feel the contours and follow the vibration to understand the spatial layout, the mutual position of the elements, their dimension, and the overall spatial rationale of the picture, following in the meantime the audio cues which explain the meaning of each contour or geometry element in the picture.

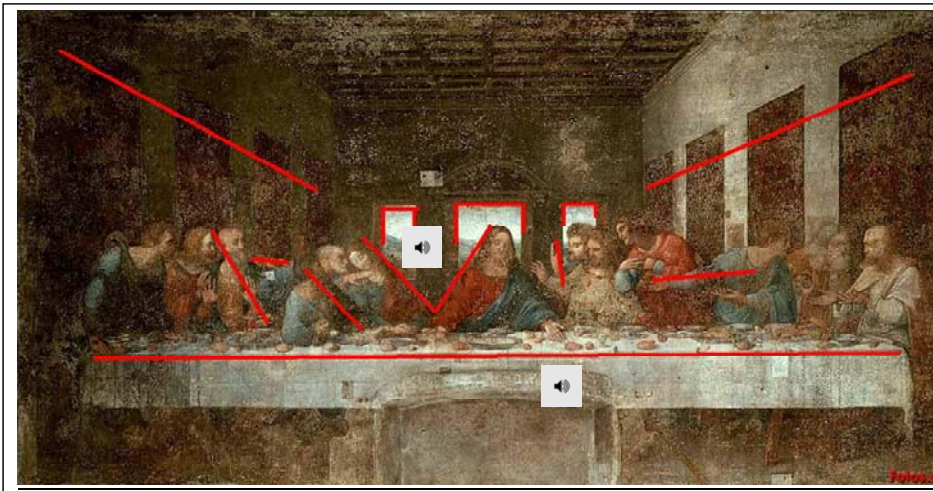


Fig. 10: The Last Supper could be downloaded via the web and made accessible to visually impaired users using vibro-aural cues that allow multimodal exploration

In this way, the textual explanation contained in the site can be corroborated by real experience – not *seeing* the picture, but *understanding* its spatial organisation and its layout by *feeling* it with other senses.

In the *Last Supper* case, for example, various elements can be shown such as: the perspective of the room by exploring the two side-diagonal perspective lines, the length of the table, the V shape between Saint James and Jesus, the mysterious, pointing finger at the right of Jesus, how the apostles are bent away from Jesus in the moment of His stunning revelation about betrayal, the straight, upward pointing finger of St.Peter, asking if he is the betrayer, etc. Each of these elements can be followed using the vibrational trail to assess its mutual position in the picture, its length, its shape, its dimension, and enriched by the audio explanation attached to it.

In future tests, we will experiment providing access to interesting paintings and trying to assess whether the interest in this subject could be a possible additional driver to capture the necessary user's concentration, provided that the content is repurposed with interesting texts, audio, and vibrational patterns.

## Acknowledgements

This project was supported under the MAIS and ITC4D National projects, financed by the Ministry of Education and University and by Regione Lombardia. Acknowledgements go to the Institute of the Blind in Milan (Franco Lisi, Francesco Cusati, Fabrizio Sordi), which helped in the design and assessment phase, and to all the students and colleagues which cooperated with ideas and contributions (Marco Ferra, Ferdinando Carella, Piergiorgio Santini, Antonio Bianchi, Barbara Domenicantonio).

*Notice: The AudioTact device and method is patent pending.*

## References

- [1] Tzvetkova-Arsova M., Model for Tactile Exploration of Objects in Small Space by Young Blind Students, ICEVI 2000, Cracow
- [2] D'Apice T., The Role of Audio Input when Presenting Tactile Graphics and Models to Children, ICEVI 2000, Cracow
- [3] T. Barbieri, A. Bianchi, L. Sbattella, Multimodal Communication for Vision and Hearing Impairments. Conference and Workshop on Assistive Technologies for Vision and Hearing Impairment (CVHI 2004), Granada (Spain).
- [4] Offen D., Thomlinson B. - Good Vibrations: Using a Tactile Mouse to Convey Page Layout Information to Visually Impaired Computer Users, CSUN2001, Los Angeles