Increasing the dimensionality of a Geographic Information System (GIS) Using Auditory Display

Ryan MacVeigh and R. Daniel Jacobson

MGIS, Department of Geography
University of Calgary
2500 University Dr. NW, Calgary, AB T2N 1N4
ryan.macveigh@ucalgary.ca

ABSTRACT

This paper describes a way to incorporate sound into a raster based classified image. Methods for determining the sound location, amplitude, type and how to create a layer to store the information are described. Hurdles are discussed and suggestions of how to overcome them are presented. As humans we rely on our senses to help us navigate the world. Sight, sound, touch, taste, and smell; they all help us perceive our environment. Although we sometimes take vision for granted, all our other senses play as important of a role in our daily lives. Even with all these senses at our disposal, the conventional GIS very uncommonly do much more than convey their information visually. We demonstrate an auditory display with a sample implementation using a classified raster image, commonly used in a GIS analysis. This was achieved using a spatial sonification algorithm initially created in a Java environment.

The ultimate aim of this work is to develop an interactive mapping technology that fully incorporates auditory display, over a variety of platforms and applications. Such a tool would have the potential be of great benefit for displaying multivariate information in complex information displays.

[Keywords: Geographic Information Systems, GIS, mapping, image sonification, auditory map display, spatial sonification, audio mapping, locational sound mapping]

1. INTRODUCTION

Although maps are still most commonly used for orientation and navigation, more information may be of interest in some cases such as annual rainfall, soil composition, pipe lines, archeological sites, etc [1]. There is only so much of this information that can be explained visually, and sometimes it is simply not enough. Sometimes, data quantities mapped to color schemes, shapes and varying textures just do not clearly describe the information to be presented as well as would be liked [2]. As well, a conventional thematic map represented by two different color schemes can be very difficult to understand, forcing the user to constantly refer to an arbitrary legend [3]. For these reasons, adding an opportunity to present more information via another sense is important.

For a number of reasons, sound is a good method of conveying information in a GIS. Sound does not obstruct visual information [1], and when used together, sound and visualization can convey information better than separately [4]. Sound patterns can be easily detected and described, regardless if the user is musically trained, and multiple observations and patterns can be detected [4][5]. As well, combining sound with a visual display may be an ideal way to deal with the increasing complexities that geographers want to obtain in their map making. Adding sound can be as viable of a means of presenting and communicating information as visual methods are [6]. Sound can increase the amount of information communicated to the user or reduce the amount of information the user is expected to perceive visually [5]. Using a crime map as an example, sound may be used to represent the distance from a police station, like a metal detector, changing the rate of change of the time between beeps. This leaves the map clear of obstructing information and available to present other information, such as crime rate, visually using a color scheme. As the user moves the mouse pointer around map the audio would change to produce the correct sound in relation to the distance.

The increasing ease of use and functionality of computers makes this task much more achievable then it would have been only a few years ago [6]. Sonification, the use of a sound generator to map data into such things as pitch, brightness and loudness, is an example [7]. The internet provides a well suited container for the incorporation of GIS and sound. Web applications provide a novel and easy to use way to merge and present all types of information, including sound.

Displaying a spatial information vision has been described as the sense ‘par excellence’, due to an observer's ability to view the permanency of the display and holistically integrate different features. There has been a long history of displaying information through well researched visual variables [8]. Similar sonic variables have been proposed for displaying auditory information within the geographic and cartographic context [6]. However, there are significant differences between the two information modalities. Information presented thorough visual channels can lead to the rapid apprehension of configurational and spatial arrangements, while information presented through auditory display is temporal and ephemeral in nature. Approaches to rendering map like visual information, through sound, have fallen into three broad categories. Firstly, those that attempt to sonify the whole of the visual display. Here positional x and y information is replaced by pitch and time [9]. This approach has worked well for simple geometric data displays such as the interpretation of line graphs [10]. With experience and training some vision impaired users have been able to use such systems in sensory substitution navigation [11]. The second major category has been to augment a visual system with the auditory information that is delivered after a query or by specifying a location from the user, such as positioning a mouse on the display. This then leads to the ability of concurrently displaying the spatial information, visually and the multivariate information through auditory display. Approaches in this area have been used in displaying uncertainty [12]. In a hybrid approach utilizing the head related transfer function (HRTF) three dimensional immersive audio has been used in an egocentric manner to display of information about the surroundings of application to illustrate where the user is. Such systems have been implemented within personal guidance systems, whereby non visual users navigate simultaneously through the real world and augmented the virtual world [13].

Auditory information has the potential to be presented in map referenced, spatialised and immersive contexts, incorporating a mixture of speech, representational sound, sonification, and earcons.

This facilitates a user to hear auditory cues that add both redundant and complementary information that would normally be presented through text labels and other cartographic depiction methods

A multimodal approach to cartographic data interaction provides novel opportunities for interacting with cartographic data. This has wide reaching relevance beyond tactile mapping incorporating innovative and original solutions to key cartographic research areas. These include, but are not limited to, environments where users may be “blind” due to bandwidth or small screen display problems in mobile computing environments, attention diversion in in-car navigation systems, or by an overloaded visual channel in complex data analysis interfaces in geovisualization and data mining interfaces, or novice users exploring unfamiliar data.

In the context of this application, sound can be broken down into two categories; abstract or realistic [6]. Vocal narration and familiar sounds (sounds that a person can relate to an event or situation) are examples of realistic sounds [6]. Abstract sounds are sounds that are created by manipulating some or all of the sounds variables to produce a sound that is meant to represent data [6]. Both have their specific purposes and depending on what data needs to be presented, no one is better than the other. Furthermore, abstract sound can be explained in a set a variables: location, loudness, pitch, register, timbre, duration, rate of change, order and attack/decay [6]. The ideas to follow focus on familiar, realistic sounds and therefore only location and loudness shall be explained and used. Loudness is the magnitude of sound, or amplitude, and location is the location of the sound in two or three dimensional space [7].

Adding audio to a visual display, in addition to adding dimensionality, can enhance the users understanding of the data as well as allow for rapid detection of information, orientating and promotes ease of learning [14]. A study performed by Lodha et al. (2000) discovered that the use of a bi-modal mapping of sound and visual data provided a more accurate understanding of the data [2].

2. METHODS

It is proposed that the sound can be added to a GIS as a layer, just as any other attribute of visual information would. Sound will be calculated for every pixel in the image using an algorithm and then added invisibly to the project. When added to the GIS, the sound layer can be accessed in real time and produce sound dependent on the data stored in relation to the location of the pixel under the mouse. As the user moves the mouse around the map, the sound will change appropriately depending on those derived values. The sounds used for this example will be realistic familiar sounds. If there is data that specifies there is water to the South, the user will hear running water from the rear channels of the stereo, if the pixels are classified as city he/she will hear traffic, etc.

It is the hypothesis of this work that using the methods discussed will create a relatively simple way of incorporating sound into a GIS to ultimately increase the dimensionality of the map being produced.

2.1. Sound Distance

Sound distance is the arbitrary distance that a human can hear if physically present in the study area or simply a distance that the map creator decides should be used to represent how far the sound can be heard. With sound, the signal inevitably degrades as it moves from the source to the point of perception [15]. The sound distance variable is the cut off point where no sound can be heard. The actual distance may be difficult to obtain but an estimation of this distance can be used since the purpose is to represent the classified image, not the real world. Once chosen, the sound distance variable is used to calculate the length of the window side using:

$$a = \sqrt{\frac{SD^2}{2}} \tag{1}$$

where a is the length of the side of the window and SD is the sound distance measured in the same units of the classified image’s spectral resolution. The result is an actual distance on the ground of the window side. This number needs to be converted to pixels using:

$$win = \frac{a}{r} \tag{2}$$

where a is the derived length of the side, r is the spectral resolution of the sensor used to record the image and win, when rounded to the nearest odd integer is the window size.

To determine the type of sound, five operations will be applied to certain locations for the point where the user is listening from (the pixel under the mouse) (Figure 1). In this example, the resulting sound will be dominated by forest sound while some city sounds can be heard in the front-left channel and some water in the rear-right channel.

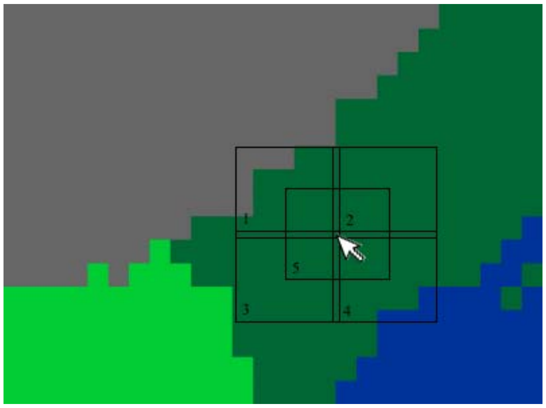

Figure 1 *The five operations applied to the pixel under the mouse to create the sound at that location*

2.2. Creating the sound layer

Once the window size has been determined the window moves across the whole image to create the sound layers. One layer will be created for each class in the classified image. In Figure 1, the classified image contains the classes forest, agriculture, city and water. The window passes four times, once counting the number of pixels in the window for forest, once for city and so on.

All the layers are then merged into one layer where each pixel in the grid will contain one value for every class. This is the layer that will be added into the GIS and used to produce the sound. The spatial extent of the layer is then set to match that of the classified image. If desired, the layers can be added separately instead of merged. How the software used to produce the sound is designed will determine which method works best.

2.3. Determining sound

Determining sound happens "on the fly". This means that the sound layer is continually accessed at the point under the pointer and sound variables will change accordingly in real- time. The pixel under the mouse is the listening point and determines the sound to be heard. Sound will be represented by five locations, the sound to the North-East, North-West, South-East, South-West, and in the direct proximity to the listening point (Figure 1). This is done to make use of the fade and balance features of quadraphonic stereo systems.

2.4. Determining sound location

The sound layer is accessed at the point under the mouse pointer to determine the sound at that location. By using the window size determined by Equation 2, each of the five center points can be found.

The center points for these five windows will be; half the window size to the North-East, half the window size to the North-West, half the window size to the South-East, half the window size to the South-West and the location directly under the mouse:

$$
\begin{aligned}
CP_{NW} &= \left(-\left\lfloor \frac{win}{2} \right\rfloor i, \left\lfloor \frac{win}{2} \right\rfloor j \right) \\
CP_{NE} &= \left(\left\lfloor \frac{win}{2} \right\rfloor i, \left\lfloor \frac{win}{2} \right\rfloor j \right) \\
CP_{SW} &= \left(-\left\lfloor \frac{win}{2} \right\rfloor i, -\left\lfloor \frac{win}{2} \right\rfloor j \right) \\
CP_{SE} &= \left(\left\lfloor \frac{win}{2} \right\rfloor i, -\left\lfloor \frac{win}{2} \right\rfloor j \right) \\
CP_{CENTER} &= (i, j)
\end{aligned}
\qquad (3)
$$

where CP is the center point, i and j are the x and y coordinates of pixels in the image and win is the window size. The division is applied the mathematical floor function to give the largest integer less than the result.

Each of these points will be used to determine the sound around the listening point. The determined sounds from the points around the mouse pointer will be mapped to different channels of the stereo. For instance, what is to be heard in the North-East will play from the left-front channel of the system. This assumes that the map has north directly up and the user assumes that the listening point is always facing north. The fade and balance is changed for each of the four directional calculations.

2.5. Determining sound amplitude

To determine the amplitudes of each of the types of sounds, the number of pixels that represent one class is divided by the total number of pixels in the window. This gives the percentage of each class in the window that can be used to determine which sounds should be louder or softer to represent the area being analyzed.

The fifth operation determines the sound in direct proximity to the listening point and is applied to and played over all the channels. This is done on the concept the sound signal degrades over distance [15] and so an anomaly further away from the listening point is not over represented while others close to the listener sound louder.

2.6. Playing the sound

When viewing the map, all sounds for each classification will be playing simultaneously. It is the value of the center points that determines the amplitude of each classification of sound. If that value is zero, then the amplitude is zero for that class. As the user moves the mouse about the screen, the amplitudes change in real time. Rolling the mouse off the extent of the sound layer will stop the sound.

3. IMPLEMENTATION

An example system has been built using the proposed algorithm. The algorithm was simplified by excluding front and rear channels by merging the front-left and rear-left channels into a single left channel. This was replicated for the right channels. The center calculations remained the same. The system was designed in Java using the functionality of the *javax.sound.sampled* package available with versions of Java later than 1.3. The map chosen for the test was a simple land cover map of an ocean shore scene consisting of three classes; land water and harbor. A feature was added to hide the map to help in determination of the effectiveness of the algorithm. Figure 2 is a snapshot of the programs interface. Land was audibly represented by chirping birds, water was represented by splashing waves and harbor represented by a ships horn. The visual and auditory items are derived from the seatouch project, which uses haptic and auditory display to provide realtime navigation information to blind sailors [16]. A set of headphones was used as speakers for testing. The sounds amplitudes and locations were calculated "on the fly" and no extra layer was created.

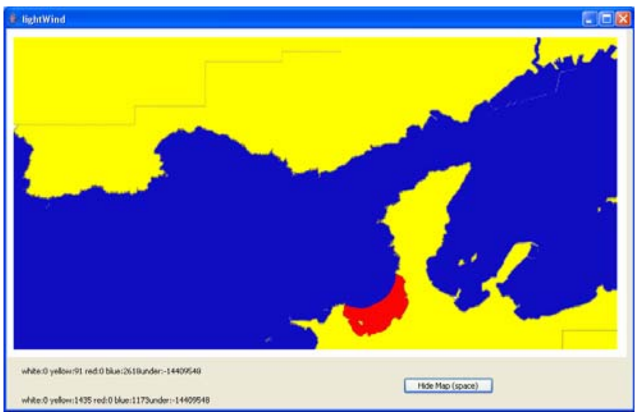

Figure 2 *The Java interface used for testing. Land is represented by yellow, water is blue and harbor is red.*

The user navigated the map using the mouse pointer and was able to easily determine what was present to the right and left of the listening point. It was found that determining size of phenomenon was more difficult but could be done by slowly moving the listening points around to find how the sound changed in the speakers. The three sounds chosen were significantly different and therefore easy to distinguish amongst even when all sounds were present. The choice of easily distinguishable sounds was important in the success of the map. For instance, the sound of wind through trees was not significantly different enough than the sound of waves and was very difficult to discern. This is analogous to visual display variables in conventional cartographic display.

Because the sound values of balance and amplitude were calculated as the user moved the listening point, there was some lag in determining the sound to play. The user needed to navigate slowly or the values would jump quickly and cause confusion. Temporal lags between user interactions and auditory display are a common problem in multimodal interfaces. Other researchers have tried to minimize these problems by maximizing CPU speed. Overcoming lag effects remains a major challenge, especially when interface areas used to represent geographic phenomena are small relative to the rest of the display. Although auditory display is excellent at alerting for time based events, in a spatial display, the triggering of audio cues from areas with small physical size may be lost due to the ephemeral nature of temporal sound display. For example, a circle representing a city in a map of a country can always be 'seen', but may only be 'heard' when triggered from its small area. In the implementation presented in this paper much of the computation time could be done before hand by the creation of a sound layer as described earlier.

Pilot testing began with only two classes; water and land. The user was quickly and easily able to decipher and reconstruct their locations. With the addition of the harbor class the user took more time to determine all of the locations when all three sounds were streamed to the speakers. This was alleviated with more practice with the program. It is clear that there is a limit to the number of classes that can be represented in this manner but what that number is and whether it varies from user to user is unclear. Further testing could be done to determine this using this implementation. This mirrors findings from Jeong and Jacobson (2002) who approximated the classic users ability to retain seven plus or minus two variables in working memory to three plus or two when using a non-visual auditory and haptic based map interface [17].

4. DISSCUSSION

Some issues need to be addressed when using sound in a GIS. Good choice and good taste must be taken into account when deciding on what sounds will be used. The study by Lodha et al. (2000) noted that many of the participants found the sounds to be annoying [2]. Just as it is important to produce visually pleasing maps, it is important to produce acoustically pleasing ones if using sound to present information. Choosing the correct sound variable to present the information is important [18]. An attempt should be made to use variables that make sense; a rise in pitch would represent a rise in temperature for example. For the user of the map to properly identify the sounds and know what they mean, a legend must be presented and the best way to do this is unclear [6]. Users may succumb to sensory overload if presented with too much information at once; this includes auditory information [6]. Care needs to be given not to present too much information at once and the use of sound in maps should be tested for this.

With a sound layer now available and a method to use that layer to determine sounds, there is the capability of use within a GIS. Unfortunately, there is no current mainstream GIS software that has this functionality. Many software suites allow a raster calculator or others forms of raster manipulation, such as convolution, which is all that is needed to create the sound layer but what is to be done with it once it is calculated?

Custom programs can be built to create and produce sounds in maps. The implementation used by the authors here is an example. Gluck (2000) has created software visualization and sonification studies that use maps specifically for spatial data analysts. Methods in such software as these can be used to create programs specifically tailored to cartographers and map users [4].

In addition to the implementation proposed in this paper, a solution may be incorporated into existing GIS software. The Environmental Systems Research Institute (ESRI) ArcGIS suite has the ability for end users to develop custom scripts and extensions that can be used within their software though Visual Basic for Applications (VBA) [19]. VBA does allow the playing of wav files and the control of volume, balance and fade by using software such as Microsoft's DirectSound [20]. It is possible that a macro can be written for ArcGIS to be able to play the sound and is worth further research. Once the map is exported it is out of the program and no longer has access to these macros. This causes a problem since now another program must be developed to not only view new map with sound imbedded but to export the format.

5. CONCLUSIONS

It is proposed that a classified image can be represented as sound in a GIS. Studies have proven that adding sound along with visual information improves the readability, the understanding and the information retention of map users [2][3][5][21]. Now that technology has advanced to a point where adding in sound to displays is much easier, GIS can benefit from the added dimensionality that the sound adds. More information can be displayed in a single map, reducing the need for multiple maps and therefore decreasing the time needed to understand and reference.

Raster datasets provide a unique way of converting their information into sound. The sound layers created are essentially raster layers whose values are represented audibly instead of visually. Although grouping of common pixels were used in the example in this paper, other operations; local, local neighborhood, extended neighborhood or zonal; may be preformed to create the sound layer. Alternatively, the sound layer may simply be the exact raster layer. What the sound is meant to represent will determine how the sound layer is created. The methods to read this sound layer can vary as well and this again depends on the type of information and sound.

The need to display more and more information is growing as cartographers try to represent more spatial information for more complex problems. Moving away from the conventional visual display to a bi-modal display with sound as well as visual information can be a way to conquer this problem and save from overwhelming the users with too much information at once.

The need for incorporating the auditory information in visual displays is becoming increasingly important, due to the complexity of information and the reduction in the screen size on portable devices. The areas where auditory information is to be of most benefit are potentially: environments where the user's visual attention needs to be directed elsewhere, such as in an in car navigation system; environments where a user becomes functionally blind, due to a hazardous environment, such as firefighters navigating in a smoke-filled building; in information systems and displays designed for partially sighted and blind people, where access to spatial information is paramount, for orientation and navigation and mobility; fourthly, in domains of complex information visualization, where the visual channel is often overloaded, due to the multivariate nature of the data.

6. REFERENCES

[1] Blattner, M.M., Papp III, A.L. & Glinert, E.P. 1994. Sonic enhancement of two-dimensional graphics displays. Auditory Display, Sonification, Audification, and Auditory Interfaces, p 447-470.

[2] Lodha, S.K., Joseph, A.J. & Renteria, J.C. 2000. Audio-visual data mapping for GIS-based data: an experimental evaluation. Proceedings of the workshop on new paradigms in information visualization and manipulation in conjunction with the eighth ACM internation conference on Information and knowledge management. p. 41-48

[3] Jeong, W. 2005 Multimodal Trivariate Thematic Maps with Auditory and Haptic Display. Proceedings 68th Annual Meeting of the American Society for Information Science and Technology (ASIST), p 42.

[4] Gluck M., 2000. The Use of Sound for Data Exploration. Bulletin of the American Society for Information Science, 26(5), p.26-28.

[5] Brewster, S.A., Wright, P.C. & Edwards, A.D.N. 1992. A detailed investigation into the effectiveness of earcons. Auditory display, sonification, audification and auditory interfaces. The Proceedings of the First International Conference on Auditory Display, p. 471-498.

[6] Krygier, J.B., 1994. Sound and Geographic Visualization. Visualization in Modern Cartography, p. 149-166.

[7] Jacobson, R. D., Kitchin, R., & Golledge, R. 2002. Multi-modal virtual reality for presenting geographic information. Virtual Reality in Geography. p 382-400.

[8] Bertin, J. 1983. Semiology of Graphics: Diagrams, Networks, Maps. Madison: University of Wisconsin Press.

[9] Meijer, P. B. L. 1992. An experimental system for auditory image representation. IEEE transactions on biomedical engineering , 112-121.

[10] Gardner, J.A , Stewart, R., Francioni, J., & Smith, A. 2002. Tiger, Agc, And Win-Triangle, Removing The Barrier To Sem Education. Proceedings of the 2002 CSUN International Conference on Technology and Persons with Disabilities

[11] Amedi, A., Bermpohl, F., Camprodon, J., Fox S., Merabet, L., Meijer, P. & Pascual-Leone A. 2005. Neural Correlates of Visual-to-Auditory Sensory Substitution in Proficient Blind Users. CNS 2005 the 12th Annual Meeting of the Cognitive Neuroscience Society

[12] Fisher, P. 1994. Hearing the reliability in classified rmotely sensed images. Cartography and Geographical Information Systems, p31-36.

[13] Loomis, J. M., Golledge, R. G. & Klatzky, R. L. 1998. Navigation system for the blind: Auditory display modes and guidance. Presence-Teleoperators and Virtual Environments, p193-203.

[14] Kramer, G. 1994. An introduction to auditory display. Auditory Display, Sonification, Audifieation, and Auditory Interfaces, p. 1-75.

[15] Naguib, M. & Wiley, H. 2001. Estimating the distance to a source of sound: mechanisms and adaptations for long-range communication. Animal Behaviour, 62, p. 825-837.

[16] Simonnet, M., Guinard, J., Tisseau, J. 2006. Preliminary work for vocal and haptic navigation software for blind sailors. International Journal Of Disabilty and Human Development. p61-67

[17] Jeong, W. and Jacobson, R.D. 2002. Exploratory user study of haptic and auditory display for multimodal information systems. In McLaughlin, M. L., Hespanha, J.P., and Sukhatme, G.S. (eds) Touch in virtual Environments:

Haptics and the design of interactive systems. IMSC Series in Multimedia. Prentice Hall, New York, pp. 194-204

[18] Walker, B. N., & Kramer, G. 1996. Mappings and metaphors in auditory displays: An experimental assessment. In Frysinger, Kramer (eds.), Proceedings of the Third International Conference on Auditory Display, ICAD '96, p. 71-74.

[19] ESRI. 2006. ArcGIS version 9.1. Redlands, California, USA.

[20] Microsoft Developers Network, 2006. DirectSound. <http://msdn.microsoft.com/library/default.asp?url=/library/en-us/directx9_c/dx9_directsound.asp>. Accessed on Dec 12, 2006.

[21] Gluck, M., L. Yu, B. Ju, W.-S. Jeong, and C.-T. Chang. 1999. Augmented seriation: usability of a visual and auditory tool for geographic pattern discovery with risk perception data. GeoComputation 99.