# Sonification: A Novel Approach towards Data mining

Zahid Halim, Dr.Rauf Baig, Shariq Bashir
Department of Computer Science
National University of Computer and Emerging Science, FAST-NU
H11/4 Islamabad, Pakistan.

**Abstract**

*The field of sonification is a subset of auditory display. It brings together interests from data mining, exploratory data analysis, human-computer interface and musical interfaces. Sonification is the mapping of data to sound; it is a rich and relatively unexplored technique for data mining. The idea behind sonification is that nonverbal sounds can be used to represent numerical data and provide support for information processing activities of many different kinds. In this paper we present three quantification rules for using sonification in data mining. We also present an RPAI (Rain Prediction Auditory Icon) Algorithm to predict rain using auditory icon. This integrates two new areas of research i.e. sonification and climate data mining. Weather data mining (forecasting) gained a lot of interest in last couple of years due to its large applications on Water, Agriculture, Energy, Health and Retail Market. The use of climate data mining is new and is in its infancy stages of being used for a variety of businesses. This creates the opportunity to explore enhancements to decision strategies through a suite of research analysis and tool development. We perform experiments using RPAI algorithm on rain data from a metrological site and discuss the results.*

*Keywords: Sonification, Data mining, Auditory icon, Visualization*

## 1. INTRODUCTION

Auditory data display denotes a rather young and rapidly evolving set of techniques also known under the term sonification to make data from a wide range of application domains accessible to auditory inspection, analysis and summarization [1]. In the context of data mining, sonification is a new concept for retrieving hidden patterns from the datasets and extracts clusters. Sonification is getting popular day by day due to certain limitations of visualization techniques for data mining which includes: *Cluster overlapping:* Most of the visualization techniques like parallel coordinates [2] and Icon based techniques suffer from cluster over lapping problems. *High dimensionality:* High dimensional data is not efficiently visualized and mined using visualization techniques. As a result congestion in the visual display increases exponentially with the increase in dimensions. *Color distinction:* Its almost impossible for human eye to distinguish between different shades of the same color like yellow and light yellow. *Limited by screen resolution:* Pixel oriented [2] and parallel coordinate techniques are limited by the screen resolution. *Limited by number of variables:* Icon based techniques work for only a limited number of attributes.

The advantages which sonification has over visualization are: Sonification creates sounds that people can listen to regarding safety data, helps people understand data since people learn in many different ways, help the blind to do their own scientific research or to learn about science data, compared to vision, hearing is non-directional, its much more difficult to miss a change in a sound.

Sonification has four techniques which include auditory icons, earcons, audification and Parameter Mapping [4].

Data mining is applied to a number of domains to extract non-trivial information, for classification and clustering. Use of data mining for climate data is in its initial stages. Given very large volumes of remote sensing data and climate model output, one would like to be able to compare them in order to understand where, when and why model data do not

agree with observations [5]. Climate data mining is useful for rain prediction and flood warning systems.

In this paper we present three quantification rules for using sonification in data mining. We also use the auditory icon technique for rain prediction. Rest of the paper is organized as section 2 describes auditory icons, section 3 presents the three quantification criterion  rules for the use of sonification in data mining, section 4 explains our RPAI algorithm for rain prediction, section 5 contains the experiments and conclusion ends the paper.

# 2. AUDITORY ICONS

We can encode information using audible events and their dimensions. The result is auditory icon. In the context of data mining we classify the dataset into sound to convey information about an event. The sound created is analogs to every day sounds producing events like crying of a baby, or sound of wind. For instance certain weather data can be classified as if it is going to rain tomorrow then the data set will produce the sound of rain. While the pitch of sound created may indicate the intensity of rain. Similarly the prediction of storm can be predicted by the intensity, pitch and rate of change of the sound produced. Auditory icons are not just for the sake of entertainment also they convey very rich information allowing us to listen to sound of data as we do to the everyday world events. These are not just for the categorization of objects or of event like those of visual icons. Parameterized auditory icons are used to reflect the relevant dimensions of the data they represent.

In the context of data mining auditory icons can provide information about user actions, possibilities of new actions and about the non-visible attributes of objects in the system. Auditory icons provide information about background process, which is especially useful in continuous process of complex systems.

## 2.1 Auditory icons Concerns

Concern regarding the construction of auditory icons, which we have identified, are:

**2.1.1 Task:** What are auditory icons good for? The basic design issue of the auditory is to identify task for the auditory icons. Although auditory icons can be used to convey very rich information but the ideal use of auditory icons is in classification.

**2.1.2 Mapping:** How is data mapped to sound? We can use the parameter of sound in the environment to the data columns to control the auditory icon. We may use the intuitive mapping between the sound and the events they indicate. Based upon these ideas mapping may be categorized into further two types.

*Intuitive*: It allows the data in the dataset to be thoughts of terms of everyday analogous.

*Appreciable*: Makes the intuitive mapping in terms of appreciable entities such as sound (icons).

Three different types of models are suggested as
- Symbolic      →      Sound of Blast
- Metaphoric   →      Musical Sound
- Iconic          →      Icon

**2.1.3 Vocabulary**: Which sound should be used for an event? How sound can both be recognized and discriminated. Hierarchal auditory icons have high vocabulary as compared to simple auditory icons.

**2.1.4 Annoyances**: How can we use sound without driving user crazy? Produce sound not annoying for users.  If you sit in a silent room you will still be listing to different sounds like shuffling of papers, footsteps of some one passing by you, sound of fan. Most of these sounds are not annoying and are useful in maintaining a background awareness of ongoing events.

## 3. CRITERION FOR SONIFICATION IN DATA MINING

Based upon the above discussion we list the following quantification criterion for the use of sonification in data mining

- *Unobtrusive:* Sonification must not drive user crazy. The sound created should be melodious. Sonification produced for data contains very sharp sound will be very repulsive and useless in the context of data mining. Sharp sounds can be replaced by relatively melodious sounds to represent those sharp sounds. Consider a sonification system being used in an emergency control room. If the sonification created for an emergency event is obtrusive it may cause people working there to panic instead of controlling the situation.

- *Informative:* Sound created should not be just music or uninformative sound. It must depict some information or represent some phenomena. Auditory icons are not just for the sake of entertainment they must convey some information allowing us to listen to sound of data. For instance the sonification for weather data should not be just pleasant sound in case they predict pleasant weather or unpleasant sounds if they predict unpleasant weather. Instead sonification for weather day should contain informative sound about different attributes of weather like speed of wind, intensity of rain or thunderstorm.

- *Variant:* Sonification for same data set should be same and different for different data sets. If we have two data sets *A* and *B* given that *A* is not same as *B*, then the sonification for *A* and *B* should be different showing that data sets *A* and *B* are different. Incase data set *A* is same as data set *B* then the sonification produce for both should depict the similarity between the two. For instance sonification for same weather condition in the month of March and April should be almost same and different weather conditions in the month of December and July should be different.

## 4. RPAI (RAIN PREDICTION AUDITORY ICON) ALGORITHM

RPAI is a probability based algorithm to predict rain through auditory icon based upon previous weather conditions. We have created a download engine that downloads daily weather report for different station in Pakistan from http://www.wunderground.com. We save the rain status of the station in a 3 dimensional vector as follows

| Station ID | Reading Date | Rain Status |
|------------|--------------|-------------|

Rain probability for the next day is calculated on the bases of rain status of previous two days as follows

Has rained today $= X$
Has not rained today $= X'$
Has rained yesterday $= Y$
Has not rained yesterday $= Y'$
Probability of rain tomorrow $= Z$

If $X$ *and* $Y$ then Z=0.8
If $X$ *and* $Y'$ then Z=0.4
If $X'$ *and* $Y$ then Z=0.3
If $X$ *and* $Y'$ then Z=0.1

After calculating rain probability we create auditory icon. We have used Infinite Impulse Response (IIR) filter, which produces an output, $y(n)$, that is the weighted sum of the current and past inputs, $x(n)$, and past outputs [6].

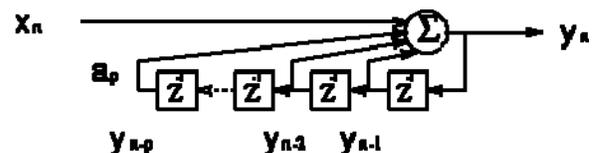The Linear Predictive model of an IIR filter is shown in Figure 1.



Figure 1 An IIR filter

Altough sampling of sound will give high quality results but we use synthesis to control parameters of auditory icon created. Sound of rain is represented as noise moving up and down with few sharp frequencies spanning a couple of microseconds showing rain drops. Thus if probability of rain and rain itself is high then the rate of change of frequency will be high and low otherwise. Figure 2 shows the affect of high and low probability on the auditory icon created.
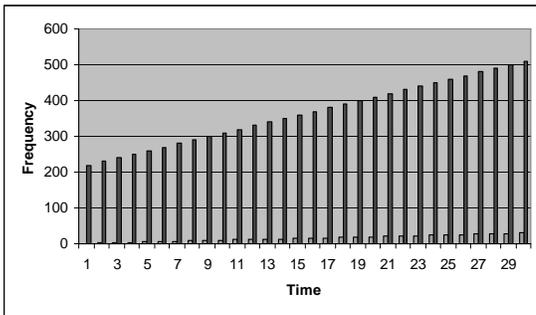


Figure 2 (a) High probability of rain shown by smooth markov chain
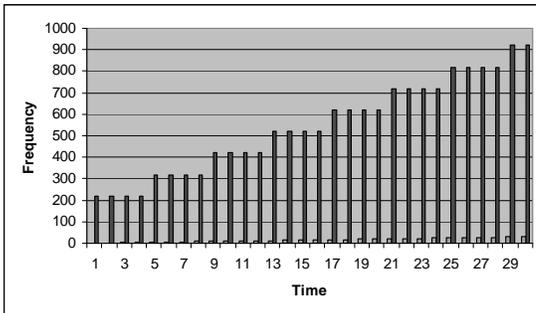


Figure 2 (b) High probability of rain shown by unsmooth markov chain

Frequencies are calculated as follows:

Frequency=150+Rnd*(Probability*10)

Time=Rnd*(150-Probability)*50

Where Probability is the probability of rain and Rnd= [0-1)

*Procedure getRainProbibility(X,X',Y,Y')*
*Start*
   *If X  and Y  then  Probability ← 0.8*
   *If X  and Y'  then  Probability ← 0.4*
   *If X'  and Y  then  Probability ← 0.3*
   *If X  and Y'  then  Probability ← 0.1*

  *Return Probability*

*End getRainProbibility*

*Procedure RPAI()*
*Start*
  *Probibility ← getRainProbibility(X,X',Y,Y')*
  *I ← 1 to 15*

       *Frequency=150+Rnd*(Probability*10)*
        *Time=Rnd*(150-Probability)*50*
        *IIF(Frequency,Time)*
*End*

## 5. EXPERIMENTS

As we mentioned above, we have developed a download engine to get daily weather updates from [7]. The URL to download weather data for each station is different. The complete URL is [8]. We have downloaded the data for the month of May and July for two stations, Peshawar and Islamabad, each with *station id 41530* and *41571* respectively. We have created a 2D plot of the data for each month and for each station. Figure 3 shows the graph for the month of May for *station id 41530*, Peshawar, showing rains status at each day of the month. The graph shows rain at $15^{th}$, $16^{th}$ and $17^{th}$ May 2006. The rain at $16^{th}$ and $17^{th}$ May 2006 was also predicted by the auditory icon created by our RPAI algorithm. Figure 4 shows the graph for the month of June for *station id 41530,* Peshawar, showing rains status at each day of the month. According to Figure 4 there was rain on $1^{st}$, $2^{nd}$, $3^{rd}$, $5^{th}$ and $6^{th}$ of June 2006 and also at $15^{th}$, $16^{th}$ and $17^{th}$ of June 2006. After running RPAI algorithm on June data for the year 2006, it created two auditory icons one showing heavy rain and the other with light rain fall.
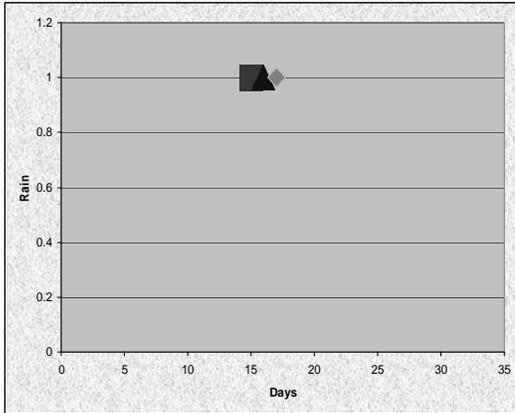
Figure 3 Rain status for the month
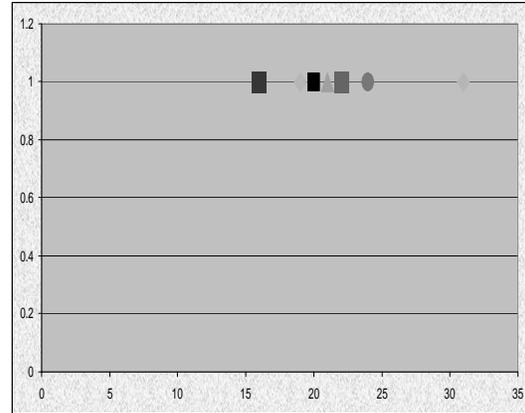of May 2006. *Station id:41530*



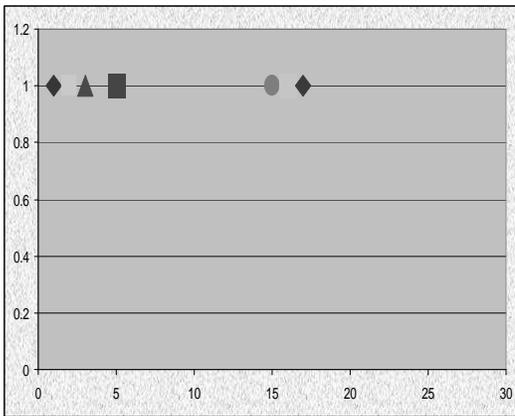Figure 5 Rain status for the month
of May 2006. *Station id:41571*



Figure 4 Rain status for the month
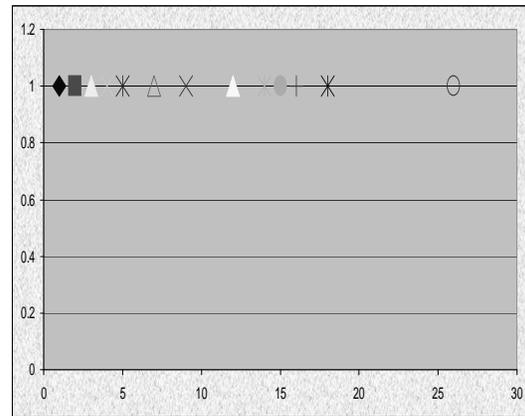of June 2006. *Station id:41530*



Figure 6 Rain status for the month
of June 2006. *Station id:41571*

In the second phase of our experiment we have tested our RPAI algorithm on the rain data of *station id 41571,* Islamabad. We have created a 2D plot of the data for the data of this station as well, shown in Figure 5 and Figure 6. Figure 5 shows the graph for the month of May for *station id 41571*, Islamabad, showing rains status at each day of the month. The graph shows rain at 16th, 19th, 20th, 21st, 22nd, 24th and 31st May 2006. The rain at these days was also predicted by the auditory icon created by our RPAI algorithm. Figure 6 shows the graph for the month of June for *station id 41571,* Islamabad, showing rains status at each day of the month. According to Figure 6 there was rain on 3rd, 4th, 5th, 9th,12th,14th,15th,16th and 18th of June 2006. After running our RPAI algorithm on the same data, it created auditory icons showing rain on these days.

It was observed that mainly four types of sound waves were produced by our RPAI algorithm. Each of the sound wave produced corresponded to one of the four probabilites of rain, 0.8, 0.4, 0.3 and 0.1 . Where sound wave corresponding to probability value 0.8 and 0.4 predicted heavy rain, sound wave corresponding to probability value 0.3 predicted low rain and sound wave corresponding to probability value 0.1 almost no rain. These sound waves are shown in Figure 7 through Figure 10.
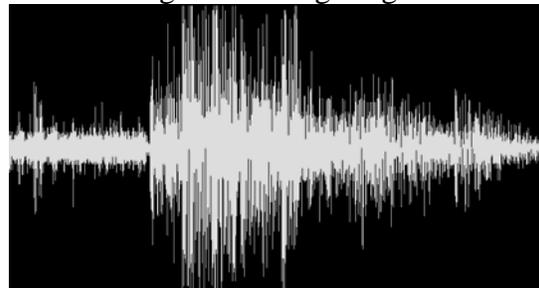


Figure 7 Sound wave corresponding
to rain probability value 0.8

Figure 8 Sound wave corresponding
to rain probability value 0.4



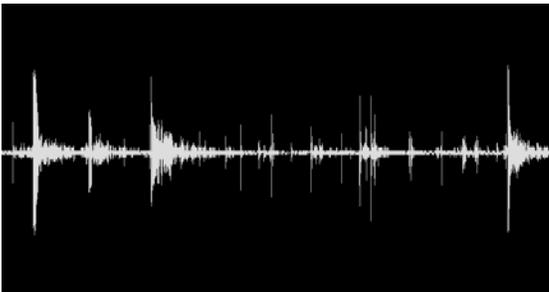Figure 9 Sound wave corresponding
to rain probability value 0.3



Figure 10 Sound wave corresponding
to rain probability value 0.1

## CONCLUSIONS

Auditory icons have a great potential as a strategy for creating informative strategy for data mining and classification. Auditory icons provide feed forward and feed backward about what must be done next.  RPAI algorithm presented here predicts the probability of rains based on weather conditions of previous 48 hours. Our RPAI algorithm can be further improved to support more weather attributes and sounds like that of thunderstorm and speed of wind. After running the RPAI algorithm on data downloaded for the two stations the auditory icons created were very accurate. Results from experiments provided rich information regarding the rain status for the next day.

**REFERENCES**

[1] T. Hermann and  H. Ritter: Sound and meaning in auditory data display, Proceedings of the IEEE, April 2004

[2] D. A. Keim: Information Visualization and Visual Data Mining, IEEE transactions on visualization and computer graphics, January 2002.
[3] D. Leeuw, V. Liere: Layout Based Visualization Techniques for Multi Dimensional Data, 2000
[4] T. Hermann, H. Ritter: Listen to your data: Model-based sonification for data analysis. ISIMADE '99, Baden-Baden, Germany, Aug. 1999
[5] R. Pincus and C. Batstone: Data Mining for Climate Model Improvement, Earth Science Technology Conference, June 2006.
[6] Infinite Impulse Response filters: http://svr-www.eng.cam.ac.uk/~ajr/SpeechAnalysis/node14.html. 2006
[7] Weather Underground: http://www.wunderground.com, 2006
[8] Weather Underground:
http://www.wunderground.com/history/station/41530/2006/5/1/DailyHistory.html?req_city=NA&req_state=NA&req_statename=NA, 2006