MONITORING REAL-TIME DATA: A SONIFICATION APPROACH

Agnieszka Roginska[1], Edward Childs[2], Micah K. Johnson[3]

[1]AuSIM Inc, 3239 El Camino Real, 2nd floor, Palo Alto CA, USA
[2]Accentus LLC, 54 Lent Rd., Sharon VT USA
[3]Department of Computer Science, Dartmouth College, Hanover NH, USA
aroginska@ausim3d.com
edward@edwardchilds.com
kimo.johnson@dartmouth.edu

ABSTRACT

This paper describes an approach to sonifying and displaying remotely sensed data. A representative sample dataset was audified using synthetic sounds, and sonified using orchestral instruments. The resulting 14 data streams were streamed in real-time and rendered using 3 display methods. Audio spatialization using HRTF processing is compared with stereo and monophonic display. Listening tests show a marked preference for the sonified data processed using HRTFs.

1. INTRODUCTION

The need to monitor remotely sensed data in real time has increased exponentially in recent years. The increase is partly due to growth in the deployment use of remote sensing devices and partly due to the complexity of societal threats such as terrorism. For example, the ECHELON global network uses input from as many as 150 satellites, in addition to microwave receivers and wide spectrum radio receivers to track voice communications for occurrences of targeted key words [1]. Other systems monitor satellite images and infrared data to track human activity in remote locations.

Aside from safety and security applications, wireless sensor networks are becoming more common in process monitoring. For example, in the oil and gas industry, 83% of the use of wireless technology is for the remote monitoring of wellheads and pipelines [2]. The technology lowers costs by reducing the need for human on-site inspections.

To date, the most common interface to the remotely sensed data is visual. The problem of “too much data” or “visual overload” is generally addressed by adding more computer screens, using sophisticated algorithms to increase the visual density of data content, immersive visualization technologies or extensive preprocessing and filtering of the data. Visual representation is limited in the number of independent dimensions of a dataset that it can simultaneously represent. In addition to spatial dimensions and a temporal dimension, a few other properties are available such as color, texture, and symbology, but these are limited in number and quickly exhausted when highly-dimensional, complex, and layered datasets are represented.

The visual system and the auditory system have complementary strengths in data display. While visual displays can provide extremely high resolution depictions of selected local areas or datasets, identifying the interesting regions for scrutiny in a very large dataset is always a challenge. Visual strategies for attacking this problem include increasing the visual display area, increasing the data display rate, coding data of greater interest with brighter colors, etc. However, even these strategies are often insufficient – a single pixel can easily be overlooked, especially in a time-varying display. Aural scanning, alternately, can be better suited for detecting subtle or transient data, as the aural sense can scan all spatial directions simultaneously without being limited to a single focal point, and can detect data features of very short temporal duration more easily and intuitively.

The use of the auditory sense in computer monitoring applications is definitely underutilized. However, the use of the ears to monitor environmental sounds (car engines, a baby waking up, odd sounds in the house, etc.) is part of everyone’s experience. Most humans develop the ability to learn and recognize sounds to a very high degree. The ability to react to even subtle changes in familiar sounds (the expectation violation) is particularly acute.

The challenge for the auditory display designer is to “invent” sounds which can be readily produced through computer speakers and which are appropriate to represent Internet traffic, financial data, barometric pressure, air pollutant concentration or any condition that must be monitored. The sounds must embody the data characteristics which are of interest to the user, and be tolerable to listen to over the course of a working day.

These techniques may also be applied successfully to the traversal or exploration of existing data sets.

1.1. Monitoring Modes

It is essential to distinguish between different monitoring modes when considering the design of auditory displays.

1.1.1. Alert Mode

Alert mode generally involves the use of a simple sound (usually stored on the computer as a fixed sound file) to indicate that a specific condition has been reached, e.g. that a patient’s blood pressure has exceeded a specified value. In the most primitive auditory systems, only one sound, such as a beep, may be available and must be used to indicate a multiplicity of conditions. In more sophisticated alerts, different sounds may apply to different conditions (or different instances of the same condition, e.g. the blood pressure for Patient A would be a bell, that for Patient B would be a gong). A further refinement would be to vary the alert sound to indicate an escalation of severity (e.g. one beep for mild b.p. elevation, two for moderate, three for severe) [3].

1.1.2. Single Track Mode

In this mode the goal is to monitor the change(s) in value of one or more data streams, e.g., financial market indicators, simultaneously. The different indicators may be distinguished using different musical instruments, and the size of movement that triggers sound (e.g. 0.5%) may be set to a high enough value by the user so that it may be heard in its own space and not interfere with the sounds of other data streams [4]. Another approach to single track monitoring is to use environmental sounds, such as rainforest, cricket or bullfrogs to indicate escalating or declining activity [5].

1.1.3. Relationship Mode

In this mode the goal is to monitor a relationship between two or more data streams, which sound simultaneously. In the case of a musical treatment, the user would listen to the counterpoint of multiple melodic lines in order to detect convergence, divergence or parallel motion. This mode was proposed for the monitoring of delta, gamma and vega in a stock option portfolio [6].

1.1.4. Global Mode

In global mode, there are generally too many data streams sounding simultaneously to be able to perceive individual streams independently. Here, the monitoring goal is to perceive the global state of a system (e.g. a computer network, or an industrial process). In the context of auditory display, one may learn to associate different system states (e.g. a "normal" state) with a specific sound. This concept, known as "auditory beacons" was pioneered and patented by Gregory Kramer [7] in 1994.

1.2. Spatial Auditory Displays

Because humans have the ability to hear in all directions concurrently, presenting listeners with spatially distributed sounds is a compelling strategy for supporting attention attraction and redirection. This allows larger quantities of data to be scanned more quickly, with attention being properly guided to the most interesting local areas of the display on a continuous basis. A method incorporating a spatial display of sonified together with a visual representation of the data may enhance the performance of the person monitoring the data to perceive more information and, in the case of critical information, may result in a faster and more accurate response. However, even without a visual display, data sonification and spatialization can be an effective standalone data presentation mechanism and result in the perception of more of the displayed information.

Having an aural display allows us to detect an "object" quickly and to continuously detect and monitor any changes in the signal. Sound has the added benefit of transporting with it the location information of the object emitting the signal. A sound occurring anywhere around a listener can be accurately localized even though it cannot be seen. Spatial information is not only important in helping the user identify the location of the originating source, spatially separated sources also have the added benefit of allowing a user to segregate multiple sources [8].

Although in recent years significant advances have been made in virtual-reality (VR) visual immersion and visual display technology, such as head-mounted displays (HMD's) and cave automatic virtual environments (CAVE's) that allow spatial traversal of high-resolution 3D data, spatial orientation can be better achieved and translational movement can be better perceived when multiple sensory systems are presented with coordinated stimuli. Data visualizations can be so abstract that a viewer can lose his or her understanding of place within the dataset, when deprived of matching kinesthetic and aural cues. Aural senses allow the cues of traversal and perspective changes to be significantly reinforced.

1.3. Literature Review

Barra, et al. [9] describe a system for monitoring work load and error conditions on Internet web servers. The system combines music chosen by the end user as background music, which is mixed with music generated from MIDI files or WAV files. In all cases, they appeared to have manipulated only fixed files. No musical synthesis or calculation of musical parameters was attempted. Hansen and Rubin [10] describe a comparable still ensure the listener will be able to perceive and understand the data that is presented.

The number of sources which can be perceived by a listener is highly dependent on many factors, including the spectral composition of the sounds, the location, the level, and others.

2. HEADPHONE SPATIALIZATION

The human auditory perception system relies heavily on binaural cues (ITD, IID, and pinna filtering) for spatial location information when it analyses the auditory scene. Binaural cues have the advantage of being independent of signal structure and can be applied to, both, pitched and unpitched sounds. It is these cues that allow us to focus in on a conversation when there are numerous speakers in one room – when binaural listening is obstructed by blocking one ear, segregation becomes more difficult [8]. Of course, a difference in spatial location alone will not cause the segregation of two sound sources, otherwise we would not be able to discern two simultaneous speakers on a monaural radio. However, spatial information might be the determining factor in a situation where two tones differ in spatial location and frequency. These two tones might be heard as one fused sound if originating from the same location.

The composite of the ITD, ILD and the spectral coloration characteristics are captured in Head-Related Transfer Functions (HRTF). Even though HRTF's are very rich in acoustic information, perceptual research shows that the auditory system is selective in the acoustic information that it uses in making judgments of the originating direction of a sound source [14]. It is also the cues contained in HRTFs which are one of the primary cues we use to segregate multiple audio streams. Due to physical differences between individuals, HRTF's vary greatly in both general shapes and detail [15][16][17][18]. As a result, serious perceptual distortions can occur while listening using HRTF's that were either synthesized or measured on another individual [18][19]. Nevertheless, research shows some individuals experience equal, sometimes improved [20][21], localization accuracy with non-individualized HRTF's – especially when HRTF's of a "good localizer" are used [14].

Spatial sound processed using HRTF's is typically displayed to a listener via headphones or loudspeakers. Loudspeaker presentation is more challenging due to the necessity of canceling the crosstalk signal [22]. Because of this, the listener is confined to a sweet spot.

A more controlled presentation of an HRTF-processed signal can be accomplished using headphone reproduction. Although this may be one of the most effective ways of presenting a binaural signal, often the problem of images perceived as coming from inside the head arises. This phenomenon occurs both with stereo and binaural signals (though less with binaural signals [23]). Adding environment cues (e.g. reverberation), using individualized HRTF's, and accounting for head motion are factors which significantly alleviate the perception of inside-the-head images.

3. DATA PROCESSING/AUDIFICATION

The dataset described in this paper has been simulated to represent remote sensing data. The dataset, consisting of 6,000,000 samples, was loaded into MATLAB and a spectrogram was computed using the short-time Fourier transform (STFT). The analysis window of the STFT was a 16384 sample Blackman-Harris window [24], the DFT size was 32768 samples, and the windows were overlapped by 75%.

Figure 1 shows a central portion of the spectrogram. The data is highly localized in both time and frequency and periodic patterns are evident in the different frequency bands. The sample rate was set to 11025 Hz so that the sounds in the highest frequency band would be around 5000 Hz.

The data was segmented into 14 different frequency bands to isolate the different periodic patterns. The bands were selected after visually inspecting the spectrogram and are shown as the unmasked (white) regions in Figure 2. The data was filtered in the frequency domain according to these bands, resynthesized, and saved into 14 separate WAV files.

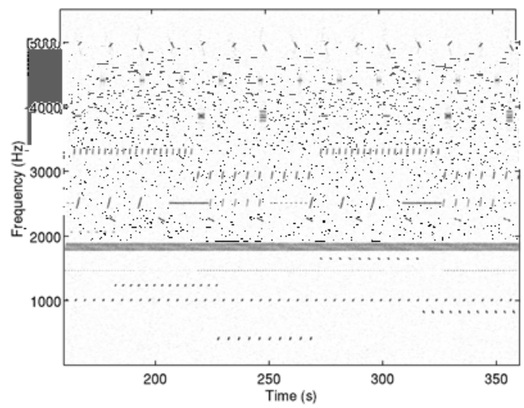

Figure 1. Spectral Plot of Sample 14-Channel Data Set.

The 14 unmasked areas of the plot were used in the sonification.

The stereo mix of these 14 WAV files, with some reverb added, may be found at: http://eamusic.dartmouth.edu/~ed/sounds/paper13_raw.mp3.

4. SONIFICATION/ORCHESTRATION

The sonification design was based on the assumption that the 14 bands identified in Section 3 represented 14 channels of real-time data to be monitored on a continuous basis. The design then sought to maximize:

1. Ease of listening over long periods of time.
2. Ability to distinguish all 14 channels.
3. Ability to perceive rhythmic, harmonic and contrapuntal interactions and relationships between the channels.
4. Perception of unusual or out of place events in the texture.

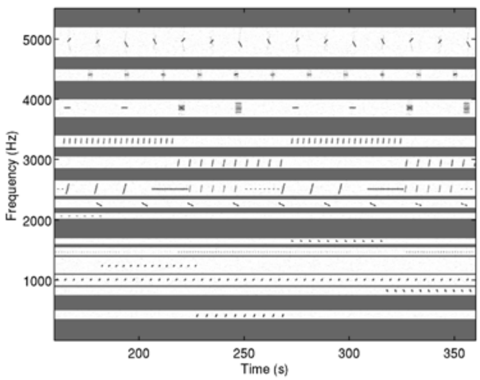

Figure 2. Spectral Plot of Sample 14-Channel Data Set.

To achieve the first objective, it was decided to orchestrate the 14 audified channels using high quality orchestral samples [25], so that the sonification could be perceived as pleasant background music. A standard classical "orchestra" was used consisting of the following instruments, pitch ranges and articulations, see Table 1.

Instrument	**Pitch Range**	**Articulation**
Flute	F#5-F#6	Legato scale
Oboe	D4-D5	Sustained chord
Clarinet	A3-C4	Sustained chord
Bassoon	D2-F#2	Staccato
French Horns (4)	F4-F#4	Slurred grace note

6. SUBJECTIVE TESTING

Subjective listening tests were presented to listeners in order to collect preliminary results on subjects' performance of identifying the multiple data streams presented to them. The goal of the listening tests were to obtain results regarding a) the number of sources subjects perceived during data signal playback and b) the identification of unusual, or abnormally active portions of the dataset.

The 14 WAV files containing the audified and sonified sounds were processed using AuSIM's 3D sound processing engine. The full audio signal length was 9m4sec. For each subject, the engine was configured to process the audio streams using one of the display methods. Sounds processed for stereo presentation used intensity panning. Spatialized sounds for the 3D audio display method were located around the listener, as described in the section above.

The subjects were seated in front of a laptop computer on which the audio processing was performed in realtime. Subjects were asked to identify the number of distinct sounds they heard during the entire course of the sound presentation, as well as indicate the times at which they detected unusual data activity from the perspective of data density and/or irregular sound event. An LCD panel was placed in front of the subject with a timer displaying the elapsed time of the sound presentation. Subjects used the displayed time on the screen to note unusual data activity.

Signal	Display	Mean	StDev
Synthetic	Mono	5.75	0.96
Synthetic	Stereo	8.5	1.29
Synthetic	3D	10.5	1.29
Instrumental	Mono	7.25	0.5
Instrumental	Stereo	10.75	0.96
Instrumental	3D	11.5	0.58

Table 2: Results of listening tests: number of perceived sources. The mean and standard deviation for the total 24 subjects, 4 subjects in each signal-display condition.

Signal	Display	Mean	StDev
Synthetic	Mono	8	1.83
Synthetic	Stereo	12.25	1.71
Synthetic	3D	15.5	1.29
Instrumental	Mono	7.25	1.71
Instrumental	Stereo	13.5	2.38
Instrumental	3D	16	1.83

Table 3: Results of listening tests: number of captivating data events. The mean and standard deviation for the total 24 subjects, 4 subjects in each signal-display condition.

A total of 24 subjects were asked to participate in the listening test. Subjects did not have any special training to listen to auditory displays. Each subject was presented with the audified or sonified data. The sound was displayed to the subject using headphones (Sennheiser HD595) using one of the three types of sound presentation: monophonic, stereo, or 3D audio. Each one of the six conditions of the sound display method and the type of sound was presented to four subjects.

Table 2 and Table 3 contain the mean and standard deviation of the results of the listening tests. Figure 3 shows a plot of the mean value of the number of perceived sources. There is a significant effect of the display method and signal type on the number of sounds perceived. The number of perceived sounds increases as the display method increases from monophonic, to stereo, to 3D audio. The figure also shows that subjects perceived a greater number of sources when instrumental sounds were used. Figure 4 shows a significant effect of the display method on the number of captivating data events. However, no significant effect is observed in the type of signal used (synthetic or instrumental).

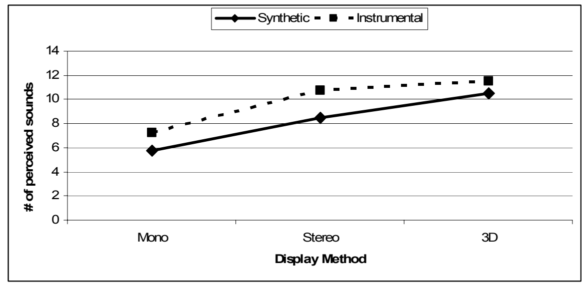

Figure 3. The mean number of perceived sound sources as a function of display method and signal type.

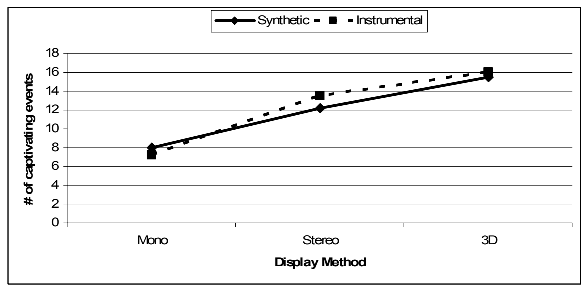

Figure 4. The mean number of captivating events detected by subjects. Results are displayed as a function of display method and signal type.

7. CONCLUSIONS

This paper presented an approach to sonifying data for the purposes of monitoring data. A dataset representative of remote sensor data was used as an example of information which may be streamed and processed for realtime monitoring. The dataset was sonified using synthetic signals and instrumental sounds. The resulting 14 WAV files representative of the 14 streams contained in the dataset were processed using three display methods: monophonic, stereo and 3D sound. Subjective listening tests indicate that data displayed using 3D sound results in a greater number of sounds perceived. This is in line with expected results based on former research indicating increased stream segregation with location cues (e.g. [8]). Sounds displayed using stereo of 3D sound show an increased number of attention-grabbing events. This indicates that a spatial presentation of signals has the capability of capturing a listener's attention more effectively. Results also show that there is a significant increase in the number of sound sources that were perceived when instrumental sounds were used to sonify the data. Typically, instrument sounds have a more familiar and characteristic spectral composition than synthetically-produced signals, and thus might result in more identifiable and distinctive sounds.

8. REFERENCES

[1] K. Coleman, "Communication Intelligence and Geographic Information in Homeland Security," *Directions Magazine, Jan. 19, 2004.* http://www.directionsmag.com/article.php?article_id=470&trv=1

[2] M. Hatler and C. Chi, *Wireless Sensor Networks for the Oil and Gas Industry,* ON World, 2005 http://www.onworld.com/wsn/oil&gas.htm

[3] D. Brock, J. A. Ballas and D. G. MacFarlane, "Encoding Urgency in Legacy Audio Alerting Systems," in *Proceedings of the 2005 International Conference on Auditory Display, Limerick, Ireland, July 6-9, 2005.*

[4] P. Janata and E. Childs, "MarketBuzz: Sonification of Real-Time Financial Data," in *Proceedings of the 2004 International Conference on Auditory Display, Sydney, Australia, July 6 – 9, 2004.*

[5] B. S. Mauney and B. N. Walker, "Creating Functional and Liveable Soundscapes for Peripheral Monitoring of Dynamic Data," in *Proceedings of the 2004 International Conference on Auditory Display, Sydney, Australia, July 6 – 9, 2004.*

[6] E. Childs, "Auditory Graphs of Real-Time Data," in *Proceedings of the 2005 International Conference on Auditory Display, Limerick, Ireland, July 6-9, 2005.*

[7] G. Kramer, "Sonification System Using Auditory Beacons for Comparison and Orientation in Data," *United States Patent No. 5,371,854,* Dec. 6, 1994.

[8] A. Bregman, *Auditory Scene Analysis*, MIT Press, 1990

[9] M. Barra, et al., "Personal Webmelody: Customized Sonification of Web Servers", in *Proceedings of the 2001 International Conference on Auditory Display, Espoo, Finland, July 29 – August 1, 2001.*

[10] M. H. Hansen and B. Rubin, "Babble Online: Applying Statistics and Design to Sonify the Internet," in *Proceedings of the 2001 International Conference on Auditory Display, Espoo, Finland, July 29 – August 1, 2001.*

[11] M. Gilfix and A. Couch, "Peep (The Network Auralizer): Monitoring Your Network With Sound," in *2000 LISA XIV, Dec. 3-8, 2000, New Orleans, LA.* http://peep.sourceforge.net/docs/lisa2000.pdf

[12] D. Brock, J. L. Stroup and J. A. Ballas, "Using Auditory Display to Manage Attention in a Dual Task Multi-Screen Environment," in *Proceedings of the 2002 International Conference on Auditory Display, Kyoto, Japan, July 2 – 5, 2002.*

[13] W.T. Fitch, G. Kramer, "Sonifying the body electric: Superiority of an auditory display over a visual display in a complex, multivariate system", In: "Auditory Display. Sonification, Audification, and Auditory Interfaces" (Ed. G. Kramer). Santa Fe Institute Studies in the Sciences of Complexity, Vol 18. Reading, MA: Addison-Wesley, 1994.

[14] E.M. Wenzel, "Acoustic Origins of Individual Differences in Sound Localization Behavior", J. Acoust. Soc. Amer., 84, 1988b.

[15] J.C. Middlebrooks, "Individual Differences in External Ear Transfer Functions Reduced by Scaling in Frequency", J. Acoust. Soc. Am., 106:1480-1492, 1999.

[16] A.W. Millis, "On the minimum audible angle," J. Acoust. Soc. Am., 100:848-856, 1958.

[17] E.A.G. Shaw, "External ear response and sound localization", in *Localization of Sound: Theory and Applications* (Amphora, Groten, CT), 30-41, 1974.

[18] E.M. Wenzel, M. Arruda, D.J. Kistler, F.L. Wightman, "Localization Using Non-individualized Head-Related Transfer Functions," J. Acoust. Soc. Amer., 94, 1993.

[19] H. Fisher, S.J. Freedman,. "The role of the pinnae in auditory localization", Journal of Auditory Research, 8:15-26, 1968.

[20] R.A. Butler, K. Belendiuk, "Spectral cues utilized in the localization of sound in the median saggital plane", J. Acoust. Soc. Am., 61, 1977.

[21] E.M. Wenzel, F. Wightman, & D. Kistler, "Localization with non-individualized virtual acoustic display cues," Proceedings of Human Factors in Computing Systems, CHI `91, pp. 351-359. New York, NY: ACM, 1991.

[22] M.R. Schroeder, B.S. Atal, "Computer Simulation of Sound Transmission in Rooms", IEEE Conv. Record, 7:150-155.

[23] G. Plenge, "On the difference between localization and lateralization", *J. Acous. Soc. Am.*, 56: 944-951, 1974.

[24] F. J. Harris, "On the use of windows for harmonic analysis with the discrete Fourier transform," *Proc. IEEE, Vol. 66, No. 1, Jan. 1978*, pp. 51-83.

[25] Vienna Sample Library, http://www.vsl.co.at/en-us/65/71/214.vsl

[26] E. Childs, "ACHORRIPSIS: A Sonification of Probability Distributions," *Proceedings of the 2002 International Conference on Auditory Display, Kyoto, Japan, July 2 – 5, 2002.*

[27] W.G. Gardener, K.D. Martin, "HRTF measurements of a KEMAR", *J. Acoust. Soc. Am.* 97, 1995.