De Natura Sonoris: Sonification of Complex Data

ODED BEN-TAL, MICHELLE DANIELS, and JONATHAN BERGER
Center for Computer Research in Music and Acoustics, Department of Music
Stanford University
540 Lasuen Mall, Stanford, CA 94305
USA
oded@ccrma.stanford.edu http://www-ccrma.stanford.edu/groups/soni

Abstract: - The graphical representation of complex and multidimensional data often falls short of affording us the clear and immediate insights that graphs offer for simpler data. We believe a sonic representation could greatly enhance our ability to interpret such data. In order to be effective, this sonification needs to be intuitive, flexible and easy to learn. Our natural ability at processing speech sounds could serve as a foundation for such representation. We describe two ways we have devised so far. The first involves mapping the data as deviation from typical vowel sounds, and the second is based on time/density coupling – both use formant filtering as the synthesis method and thus can be used simultaneously. We applied these methods to two sets of data: meteorological readings from the Mediterranean Sea, and stock market data. We believe these methods could yield beneficial applications in these fields as well as in the interpretation of medical, demographic, or meteorological data.

Keywords: - Sonification, Formant Synthesis, Vowels, Data representation.

1 Introduction

Graphical representation of data is an efficient and proven means of conveying essential information, but what happens when the data is very complex and/or multi-dimensional? With data such as meteorological data, demographics, some medical data, or stock market data, visualization alone is often inadequate. In such situations, it is possible to interpret the information by representing it with simplified versions or subsets (such as slices of a 3D image). However, using an auditory display to replace or augment a visual representation could be a more effective way of rendering the data.

The common human activity of listening to music usually involves the listener following multiple strands simultaneously. Moreover, each individual sound that the listener hears is constructed from a combination of multiple varying aspects. Our ear's natural ability to track these types of changes could be used to render interdependent variables onto a sound in such a way that a listener could gain important insight into the meaning of the represented data.

2 The Sonification Process

To be capable of offering insight, an auditory representation must be intuitive, easy to learn, flexible, and capable of depicting multidimensional data. We believe that an effective step towards such a goal is the utilization of synthesized vowel sounds. Sir Isaac Newton once observed that he could hear vowel-like sounds while pouring beer into a glass, and since his time, considerable cognitive research has proven that humans are highly skilled at identifying vowels and categorizing timbre according to vowel matching [1][2].

2.1 Formant Filters and Vowels

Inspired by Newton, we decided to use subtractive synthesis based on formant filters to create vocal-like sounds. In this method a set of 3-5 filters whose center frequency, bandwidth, and gain correspond to typical vowel formant values are used to filter either white noise (non-pitched) or a pulse-train (pitched). Our experiments were carried out using the Common Lisp Music synthesis language [6] to perform the filtering operations. We created flexible code that allowed us to synthesize various vowels, transition between vowels, and modify any filter settings as necessary.

The first data that we examined consisted of oceanographic measurements in the Mediterranean Sea (provided by Artur Hecht, Israel Oceanographic & Limnological Research) - temperature and salinity at varying depths of up to 100 meters in 5 locations. We mapped the temperature and salinity values onto the frequency and bandwidth of formant filters such that the average of the data corresponded to the typical value for a vowel. Sequential changes to filter frequency and bandwidth over time corresponded to increasing depth. We used both a continuous sonification and discreet bursts, with the latter proving more effective. Using this basic mapping, we could hear similarities and differences between the data from the five different locations and could also identify instances of atypical behavior. The sonic patterns we produced also proved relatively easy to remember in a similar way to recalling a melody from a song. This is quite unlike a jagged line in a graph, whose details are very difficult to memorize.

These were encouraging results, but in order to achieve our goals we needed to sonify multiple sets of data. This would entail using simultaneous vowel streams which the listener would be required to separate and follow.

2.2 Concurrent vowels

Concurrent vowel identification is a fairly new area of research, with the most revealing work completed only in the 1990's. Research in this area is focused on pairs of vowels, and findings mostly concentrate on three attributes of these pairs: a difference in the fundamental frequency of the vowels [3], the duration of the vowel [4], and the effects of formant transitions in one of the vowels [5].

Taking advantage of differences in fundamental frequency between vowels in our sonification process would force the use of a pitched sound source (such as a pulse-train) for filtering. Unfortunately, bandwidth changes are much less audible using the pulse-train than the random noise source, and they draw the listener's ear to focus more on the pitch of the sound rather than on timbral changes. Because of these drawbacks, we concentrated our efforts primarily on the other two attributes. Our ability to differentiate concurrent vowels improved slightly for longer bursts with durations of up to approximately 0.5 seconds but did not change much with further increases.

Because our aim is to represent changing data, the presence of formant transitions is built into our model. A significant improvement in our ability to separate vowels occurs when one vowel is constant while the formants of the other are changing. Thus some differences in behavior between two sets rendered simultaneously will allow one to become salient. This trait can be utilized when these differences are the target sought by a user. We can also introduce relatively small changes in the filter setting of one of the vowels while keeping the other settings static (for each burst of sound representing one simultaneous point). This is somewhat akin to contrast enhancement for images. Using staggered bursts, with some microchanges we were able to map two data sets simultaneously.

2.3 Density

Since time is one of the most important aspects of music. Could time and the perception of time be utilized more directly for sonification as well? This could also have the advantage of allowing us to expand our ability to represent multidimensional data in parallel to the methods we have been exploring thus far.

Using time as a varying parameter by mapping one variable to the duration of a sound burst (still using filtered noise) and another variable to the time interval between the bursts, we get a density-based representation of data. Long time intervals yield a pulsed pattern gradually giving way to denser textures with shorter intervals. This is modified by variation in the duration of each pulse. Thus this representation contains the possibility of a coupling of the variables to create an emergent totality. Furthermore, it contains an inherent staggering of onset times, which, as mentioned earlier, increases a listener's ability to segregate concurrent vowel presentations.

We implemented this mapping using stock market data. Each day of trade was represented by one second of sound. The closing price of the day was mapped to the number of bursts, while the volume of trade was mapped to the duration of each burst. Thus high volume, high price days were heard as loud, dense sounds, as opposed to sounds with pulsed rhythms for low volume, low price days. Various degrees of loudness and speed afforded a gradation of intermediate states. We could also sonify two stocks simultaneously using two different vowels.

Our preliminary studies also show that it is possible to overlay these density parameters over the filter-setting changes discussed earlier to produce a 4-dimensional representation, but further studies are needed to refine this method.

3 Tools for Sonification

To make our existing programming-based tools more useful and user-friendly we are developing a full-scale application whose graphical user interface is designed with the goal of enabling a user to experiment in real-time with the sound of data. This application, written in C++ using the Synthesis ToolKit [7], will allow a user to input multiple sets of data, perform basic linear manipulations on them, and interactively map them to the sound parameters we are using. These capabilities offer us the opportunity to efficiently explore the possibilities of various methods and carefully refine the parameters used for scaling and mapping the data. They will also enable a naïve user to implement a sonic representation of a data set without the need for a thorough knowledge of digital filters and vowel formant synthesis.

4 Future

Currently, we are working on two fronts in our research. One goal is to improve the graphical user interface in order to make it more flexible and user-friendly. In addition, we are developing ways for the application to accept various formats of data input and allow multiple options of sonic representation with parametric control.

On the synthesis side we are working to improve our ability to render multiple vowels, and we are in the process of fine-tuning both the density-based and timbral-based representations along with their conjunctions. We are also looking at ways of generalizing the sonification process by attempting to find common methods applicable to a wide range of data formats and usages that could help serve as guidelines for using these sonification methods.

In the future, we hope that sonification will offer new ways of comprehending data of various sorts while being used by scientists, doctors, stockbrokers or others in much the same way as graphs and charts are in use for a multitude of purposes today.

References:

[1] Assmann, Peter F., Nearey Terrance M., and Hogan John T. “Vowel identification: Orthographic, perceptual, and acoustic aspects” *Journal of the Acoustical Society of America* vol. 71 no. 4 (April 1982): pp 975-989

[2] Bregman, Albert S. *Auditory Scene Analysis* MIT Press. 1999. See chapter 6 for overview of speech perception.

[3] Assmann, Peter F., and D. Dwayne Paschall. “Pitches of Concurrent Vowels.” *Journal of the Acoustical Society of America* vol. 103 no. 2 (February 1998): pp1150-1160.

[4] McKeown, J. Denis, and Roy D. Patterson. “The time course of auditory segregation: Concurrent vowels that vary in duration.” *Journal of the Acoustical Society of America* vol. 98 no. 4 (October 1995): pp 1866-1877.

[5] Assmann, Peter F. “Modeling the perception of concurrent vowels: Role of formant transitions.” *Journal of the Acoustical Society of America* vol. 100 no. 2 (August 1996): pp 1141-1152.

[6] http://www-ccrma.stanford.edu/software/clm

[7] http://www-ccrma.stanford.edu/software/stk