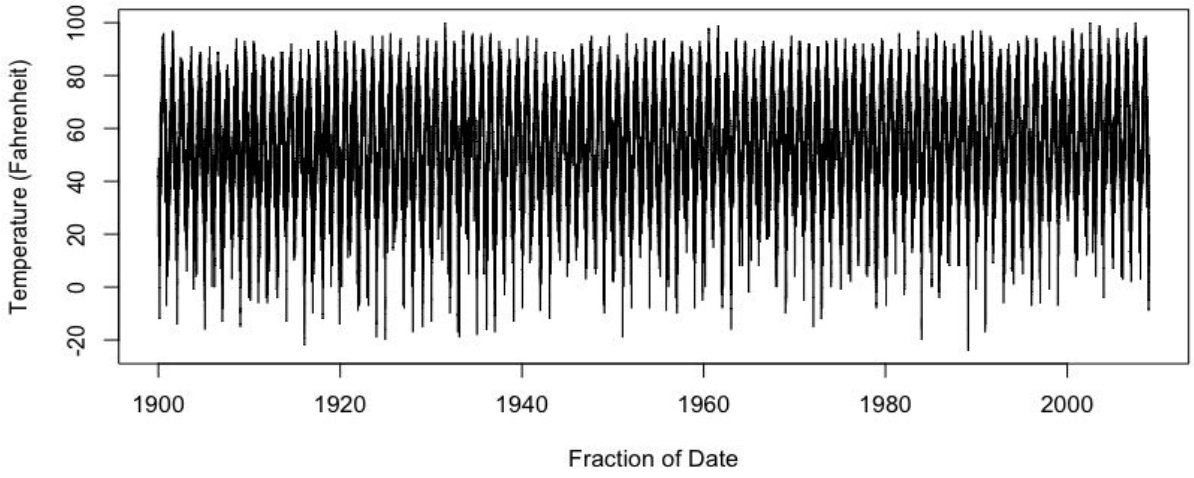
Algorithms for Climate Data Sonification

Huafeng Zhang

Introduction

Data visualization is a powerful tool in analyzing and presenting data, especially big data. Historically, researchers used bar charts, line plots, and histograms both to aid analysis and to communicate their analysis with lay people. But people with visual impairments cannot use graphs to understand their data. Even for people with perfect vision, these tools are inadequate: using them, it may be hard to recognize patterns (such as groupings and trends), anomalies, or relationships between variables in complex and large datasets such as those studied in climate research, which is more and more common in the big data 21st century.

For example, Figure 1 displays daily maximum temperatures in Bozeman, and although the plot shows only two variables (temperature and date), it is hard to identify whether there are changes in seasonal climate and long term trends from this plot. Climate science is one example of a field that grapples daily with the technological demands of analyzing and displaying such complex data sets so as to reveal as much information as possible to both scientific and nonscientific audiences.



(Figure 1: Bozeman Temperature from 1900 to 2008)

It is time to explore data presentation tools that will address the basic limitations of visualization methods. One promising direction is data sonification. Our ears can be as useful as our eyes in data analysis because of "the known superiority of the hearing sense to discriminate certain kinds of structures."[1]

In this project, I will explore several methods for producing sonification of climate data, and test how sonification can enhance users' ability to understand climate change. First I will create both sonification and visualization for climate data such as the Bozeman daily temperature data from 1900 to 2008.[2] Then I will share these files with my sponsors, statistics faculty and my colleagues at MSU Statistical Consulting and Research Services, and get feedback from them. Using their feedback, I will modify my sonification and write a report about my findings. My hypothesis is that sonification makes some features of the data more readily apparent than does conventional visualization, as measured by the feedback from statistics faculty and researchers.

Background

Data sonification is the use of non-speech sound to represent data. The Geiger counter, invented in 1908, has been used for over 100 years, and represents one example of sonification.[3] Data sonification plays a crucial role not only in helping people with visual impairments to "see" their data with their ears, but also in helping people with perfect vision present more information, identify patterns, and characterize interesting phenomena from multivariate data. Furthermore, sonification doesn't require visual attention, and people can hear the data while their hands and eyes are busy with other tasks.

Researchers are currently applying sonification to analyze time series data. McIntosh et al., (2013) stressed the importance of adding the data sonification approach to "the empirical economists' toolkit for interpreting economic data and specifying econometric models."[4] Thomas Levine (2015) used R to create a music video for financial crisis data.[5] Like McIntosh et al., the climate data that I will use in this

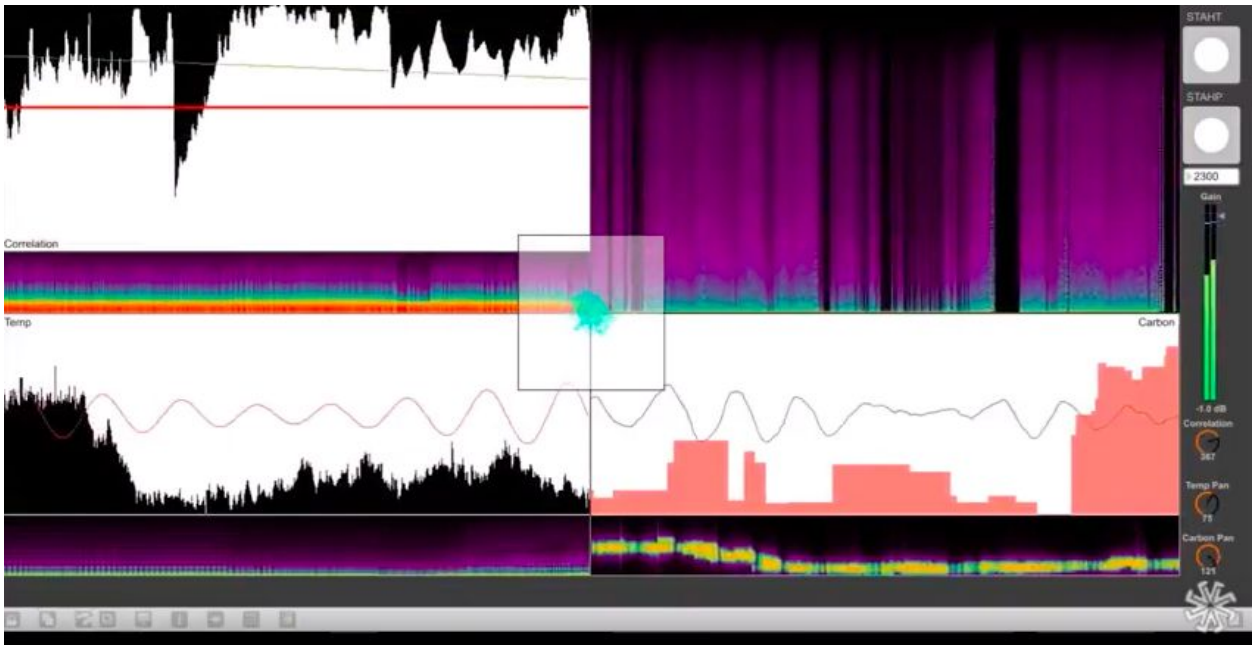
research are also time series; like Levine, I will use R and RStudio to generate the results, as they are among the most popular and powerful statistical programming tools available.[6]

My experience and education prepare me well for investigating and modifying algorithms for sonification of data. Currently, I am a senior majoring in Mathematical Sciences with a focus on Statistics and a minor in Economics. Three courses in particular--*Masterworks in Music, Introduction to Applied Multivariate Analysis, Statistical Computing and Graphical Analysis*--have given me a basic background in music, and taught me how to analyze and display multivariate data in R. Doing this project can help me build more skills in data presentation, which is crucial to my career goal, to be a professional data analyst. It also gives me a chance to collaborate with and learn from data analysis experts such as my sponsor Dr. Greenwood. Both skills will play an important role in aiding me to achieve my academic and career goals.

Methods

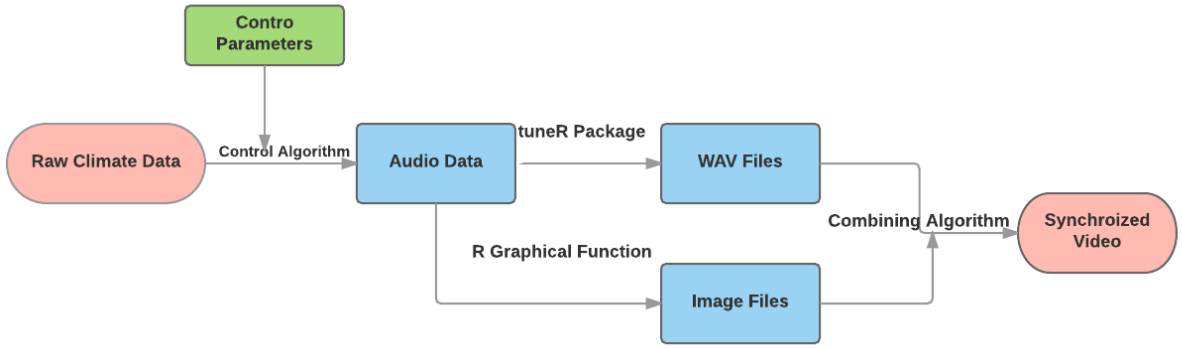
To accomplish this project successfully, my first step will be to use the programming tools R and RStudio to explore and clean the data to prepare for further data manipulation and analysis. After defining sound parameters such as rhythm, tempo, pitch, volume and so on, I will map the climate data into those parameters based on the information that I want to present, such as variations of seasonal temperature patterns, long term temperature trend(s), and single "unusual" temperature(s). Conventional graphical visualizations will also be produced based on the same data, both as static charts which must represent time along some graphical dimension, or dynamic images which represent time dependence directly.

In order to create audio and graph files, I will apply packages such as tuneR[7] and ggplot2,[8] as well as write my own functions in R. Finally, I will combine my audio and graph files to produce synchronised video representing the data in R. With these videos, I will create an integrated multichannel video that allows people to select one or more channels representing the information they want to get from the climate data. The video surface might be similar to Figure 2:[9]



(Figure 2: Sonification of 400,000 years of climate data, by Kyle Vanderburg,[9] 2015)

Having created the graph and sonification files of the climate data, I will share my videos with my sponsors, statistics faculty and my colleagues in SCRS. Then I will modify my project based on their feedback. The main steps of creating sonification for this project can be seen from the flowchart below:



(Figure 3: Steps to Create Sonification in R)

Timeline

Week 1 (5/8/17 - 5/12/17)	● Explore and clean the climate data with R.
Week 2 (5/15/17 - 5/19/17)	● Define the sound parameters of the data; ● Test how clearly these sounds represent key features; adjust as necessary.
Week 3 (5/22/17 - 5/26/17)	● Produce image files of the data with ggplot2 package in R.
Week 4 (5/29/17 - 6/2/17)	● Produce WAV file with tuneR package & other algorithm in R.
Week 5 (6/5/17 - 6/9/17)	● Produce synchronized audio and video.
Week 6 (6/12/17 - 6/16/17)	● Share the videos with my sponsors and statistics faculty.
Week 7(6/19/17 - 6/23/17)	● Collect and evaluate their feedback.
Week 8 (6/26/17 - 6/30/17)	● Modify sonification files with Dr. Greenwood.
Week 9 (7/10/17 - 7/14/17)	● Finish writing the report of the project.
Week 10 (7/17/17 -7/21/17)	● Prepare and give final presentation.

Collaboration with Faculty Sponsor

Dr. Greenwood, Associate Professor of Statistics, is my major advisor as well as my teacher in two 400 level statistics courses. He helped me clarify my idea of data sonification for time series data, and since he works with scientists at MSU familiar with climate data, he has access to relevant data sets. We will continue to discuss this project during the spring semester. He also introduced me to Dr. Hancock, Assistant Professor of Statistics, who has a background in both statistics and music. Dr. Hancock will help with the use of music in this project, for instance by defining sound parameters such as rhythm, tempo, pitch and volume for the climate data. In order to keep tracking my progress, Dr. Greenwood and I will have weekly meetings.

References

[1] Worrell, D. (2011) An Overview of Sonification. In Dean, R. T. (Ed.. *The Oxford handbook of computer music.* New York: Oxford University Press.

[2] Laboratory, O. R. (2014). Long-Term Daily and Monthly Climate Records from Stations Across the Contiguous United States. Retrieved February 13, 2017, from http://cdiac.ornl.gov/epubs/ndp/ushcn/ushcn.html

[3] Cryer, J. and Chan, K. (2008) Time series analysis: with applications in R, Springer.

[4] Kuchera-Morin, J. (2009). Transcript of "Stunning data visualization in the AlloSphere". Retrieved February 12, 2017, from https://www.ted.com/talks/joann_kuchera_morin_tours_the_allosphere/transcript?language=en

[5] Levine, T. (2015). Plotting data as music videos in R. Retrieved February 12, 2017, from https://thomaslevine.com/!/user-2015/#synthesizing-music-with-r

[6] Stanton, J. (2015). Sensing Big Data. *Big Data at Work: The Data Science Revolution and Organizational Psychology*, 158.

[7] Uwe Ligges, Sebastian Krey, Olaf Mersmann, and Sarah Schnackenberg (2016). tuneR: Analysis of music. URL:http://r-forge.r-project.org/projects/tuner/.

[8] H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2009.

[9]K. (2015, April 07). Cascades: Sonification of 400,000 years of climate data. Retrieved February 14, 2017, from https://www.youtube.com/watch?v=POMtzGV9ta8