Task 3: Preservation planning

Asztrik Bakos (1527997)
e1527997@student.tuwien.ac.at

Contents

Project application phase

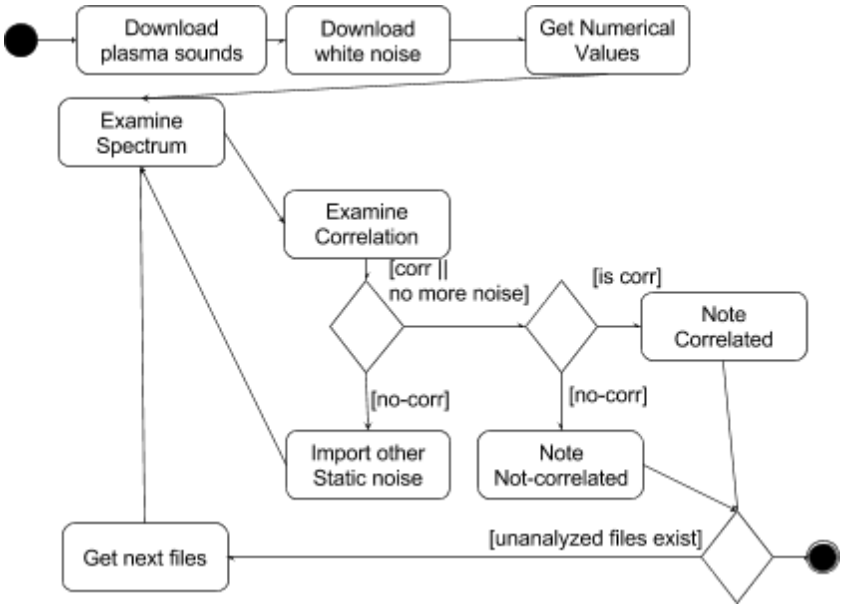
Experiment Overview

Sounds in Space

The Voyager 1 and 2 spacecrafts have travelled many billion kilometres since their launch, and sent back some very interesting data too [1]. Both Voyager 1 and 2 are equipped with the Plasma Wave Subsystem (PWS) which detects Electrical field components of plasma waves in frequency range of 10 Hz to 56 kHz. The collected data was sent back to Earth, and the plasma waves in the 20-20000 Hz range were converted into audio and made public on Youtube [2] and Soundcloud [3].

Many people believe (see comments under the Youtube-videos), that these mysterious sounds are messages of unknown civilizations, others just find it relaxing and interesting without any meaning. In my experiment I want to compare the sounds using correlation [4] to a common *white noise* like static radio noise or rainfall. Aim is to prove or disprove the randomness of the plasma waves.

Experiment UML activity diagram



Mapping:
Download: Nasa API, wget
Preprocess: ffmpeg, sox, sed
Examine: LibreOffice Calc

Subtasks

To avoid very long explanations about ownership and copyrights, it is assumed, that the sound files were either already present on my hard drive or downloaded from a webserver legally with wget.

The “Get Numerical Values” bubble contains the whole preprocessing done by a script:

```
echo "Converting $1"
ffmpeg -i $1 -acodec pcm_u8 -ar 22050 $1.wav
echo "Conversion done. Now get text"
sox $1.wav $1.dat
echo "Text written, do CSV"
sed 's/\s\+/,/gw out' $1.dat
echo "Done."
```

- **ffmpeg** converts the mp3 files to waveform audio files, because this format stores directly the time/amplitude vaules of the sound waves.
- **sox** (http://sox.sourceforge.net/) is a freeware tool capable to convert a wav file to numerical time/amplitude values and save to a textfile
- **sed** is needed because the output from sox is fixed-length and so hard to process further. To make the experiment easier, a csv format is created by some replaces.

Other publications

Since the data origin is the NASA, the plasma waves were analyzed, and several publications are available [5] [6]. But none of them answers the question whether these waves are encoded messages or not. The papers focus on physical effects of the planets or the Sun, and search for peaks and outliers in the data, while my focus is the constant roaring which also might be a messages according to some assumptions. On the other hand I want to prove or disprove, that background noise originated from an often-repeating source are just as random in the space in form of plasma waves as are on Earth when for example the rain falls.

Project execution phase

Execution monitoring and model analysis

- First I executed the experiment, by running the maker.sh script.
    - It generated the
        - Wav
        - Dat
        - Csv files
    - Then I summed up the results in the Spreadsheet
- Then I issued

```
java -jar capture-1.0-SNAPSHOT.jar -a ~/Documents/TUW/2DP/3/maker.sh
-d ~/Documents/TUW/2DP/3/
```

  Which started the capture, and lasted about two minutes.
- I verified the results by looking at the /**repo** folder, which was populated with the experiment files.
    - Also the **strace logs** found the **maker.sh**
- Finally I opened Protége, browsed the entities and wrote the SPARQL queries.

SPARQL queries

Included in sparql.txt along with the results.

Data Characterisation

I used FITS and C3PO, because the installation of Droid took some time and resulted always in errors. On the other hand, these tool include some Droid features, and result the same output.

Total number of files

71

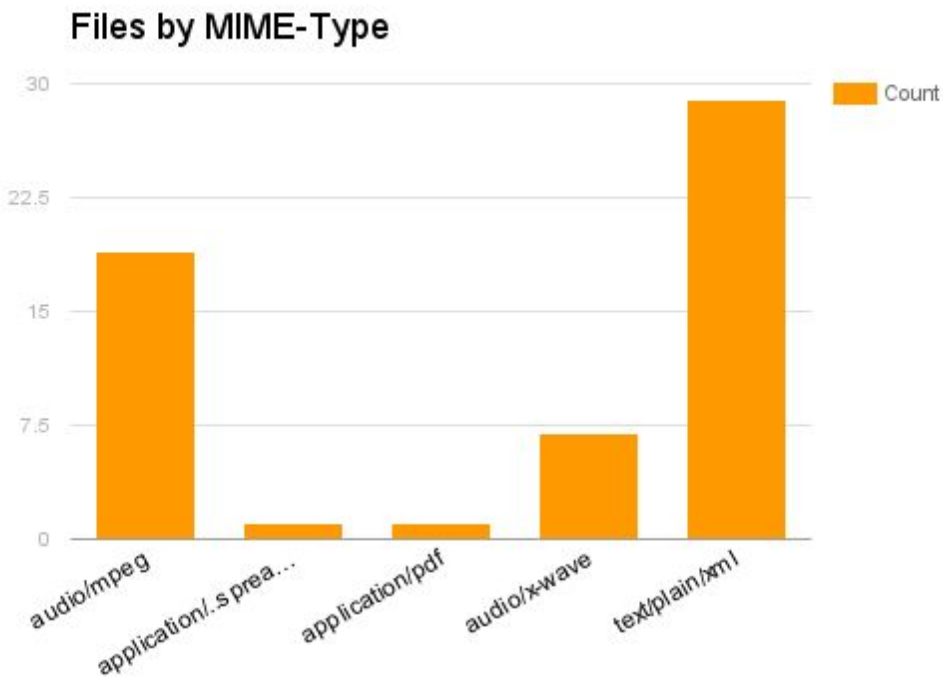
Total collection size:

2 GB

Average Filesize

29.7 MB

The MIME-Types are as follows:



Validation requirements

As identified in the previous section the experiment needs audio files and returns text. The question is whether the two waveforms correlate. To be able to decide that, the following requirements must be fulfilled:

#	Requirement	Metric	Target Value	Measurement	Tool
1	Audio conversions resulting in the same numerical waveform values. The MP3 → WAV conversion is less important, the identity of the sound can be "manually" checked. Important is to preserve the WAV → DAT function	Correlation between two DAT Value columns generated from the same WAV	1	DAT files	Spread sheet
2	Audio files must be transformed to an identifiable Waveform audio	Mean Opinion Score scale → humans have listen and rate	Over 3.5	WAV file played	Audio player
3	DAT files should be processed by a software which can calculate correlation on large datasets	Is it convertible to CSV (table form?) yes/no	yes	Import into spreadsheet software	Libre Office Calc

DMP Creation

The DMP creation was easy, and answering the questions made again clear the purpose of the project and helped to improve it by making clear that the audio files which are downloadable anyway don't have to be part of any backup.

Also the web interface is well useable, and the progress bar above helps a lot psychologically.

Sharing

The experiment files are published here:

https://github.com/asztrik/spacesounds

References

[1] http://voyager.jpl.nasa.gov
[2] https://www.youtube.com/playlist?list=PLKzoKueiKNwtP6wJF0Qa5Pn8EAtReMtcQ
[3] https://soundcloud.com/nasa/sets/solar-system-beyond-sounds
[4] https://www.mathsisfun.com/data/correlation.html
[5] http://science.nasa.gov/science-news/science-at-nasa/2013/01nov_ismsounds/]
[6] http://www-pw.physics.uiowa.edu/voyager/v1pws_interstellar_2014.html