Tampere University

Environmental audio analysis

Student: Anh Huy Bui

ID: 293257

Table of Contents

Table of Figures

1. Introduction

- This project purpose is to practice analyzing audio characteristics.
- There are 3 main stages of project: audio annotation, audio analysis and comparison and conclusion.
- Method of analysis used in this project:
  + Audio feature extraction: MFCCs, Chroma.
  + Similarity calculation: Cosine similarity, Dynamic time warping.

2. Data annotation process

2.1. Describe the annotation process

Each student was provided with 131 audio samples to be annotated and classified into different sound categories such as “adults talking”, “children voices”, “traffic noise”, etc.

- Good:
- From my point of view, the annotation UI was easy to work with, also come with a progress bar for tracking.
- Many different audio signals to be annotated.
- Bad:
- The audio cannot be paused, played at a specific time.
- Some samples have noisy front ground (of wind noise maybe) while main sound (people chatting) is in the background. This can affect analysis result.
- Possible improvement:
- Add feature that allow annotator to pause the audio at some points or to start at some points.
- Allow annotator to add categories such as “rain”. In the end, all the audios which are in the same class will be analyzed together, so it may be easier to compare audios such as those with noisy front ground of wind or rain because they are in the same class.

2.2. Dataset Statistic

From my annotation result, I divided the audios into 6 classes including: “adults talking”, “birds singing”, “children voices”, “footsteps”, “siren” and “traffic noise”. There are also some sub-classes which are a mixtures of the major ones.

Please take a look at my *.xlsx. It is easier and more informative. I also arrange and rename my annotation data according to this.

H293257.xlsx

3. Audio Analysis

3.1. Implementation

All parameters as default value such as: n_fft = 2048, hop_length = 512, n_mels = 128. I believe there is trade-off between accuracy and program size/memory cost + calculation time. So default value may give me balanced factors.

The parameter that I tried to change was n_mfcc = 40 and 80. However, the results does not vary much.

I implemented both methods including cosine similarity and DTW distance. However, in this report, I would mostly focus on Cosine Similarity.

3.2. Results and discussion

3.2.1. Heatmap and average similarity

Class	All	1					2
Sub-class	-	1.1	1.2	1.3	1.4	1.5	2.1
Description	-	Adults talking	Adults talking+children voices	Adults talking+ footsteps	Adults talking+ Birds singing	Adults talking+ Traffic noise	Children voices
Average similarity	0.99594	0.99664	0.99544	0.99751	0.99664	0.99715	0.99825

Class	3		4		5		6	7
Sub-class	3.1	3.2	4.1	4.2	5.1	5.2	6.1	7.1
Description	Birds singing	Birds singing + Traffic noise/ footsteps	Traffic noise	Traffic noise + footsteps	Footsteps	Footsteps + Other sound	Siren + other sound	Uncleared
Average similarity	0.99853	0.99772	0.99454	0.99695	0.99744	0.99949	0.99510	0.99600

3.2.2. Report from result data

- By arranging audios from same class nearby each other, heat map graph has fairly bright pixels (high level of similarity) along the white sanity check line.

- I can observe that within the same class, the first sub-class, which I marked with only one pure sound, often has higher level of similarity compared to other sub-classes.

- The results vary little (from 0.975 to 1). This can be because the collected audios are random (recorded at random time, in real places) which are not typical for each class, so differences may be little. However, in class 1, “adults talking”, similarity levels wildly fluctuate compared to others. In my opinion, people voices are more diverse than other sound such as birds, footsteps or siren, which are more typical.

- I have also tried with DTW distance and got the same result (the levels are inverted because distance and similarity are invert of each other). Please check attached figures for better resolution and explicit data.

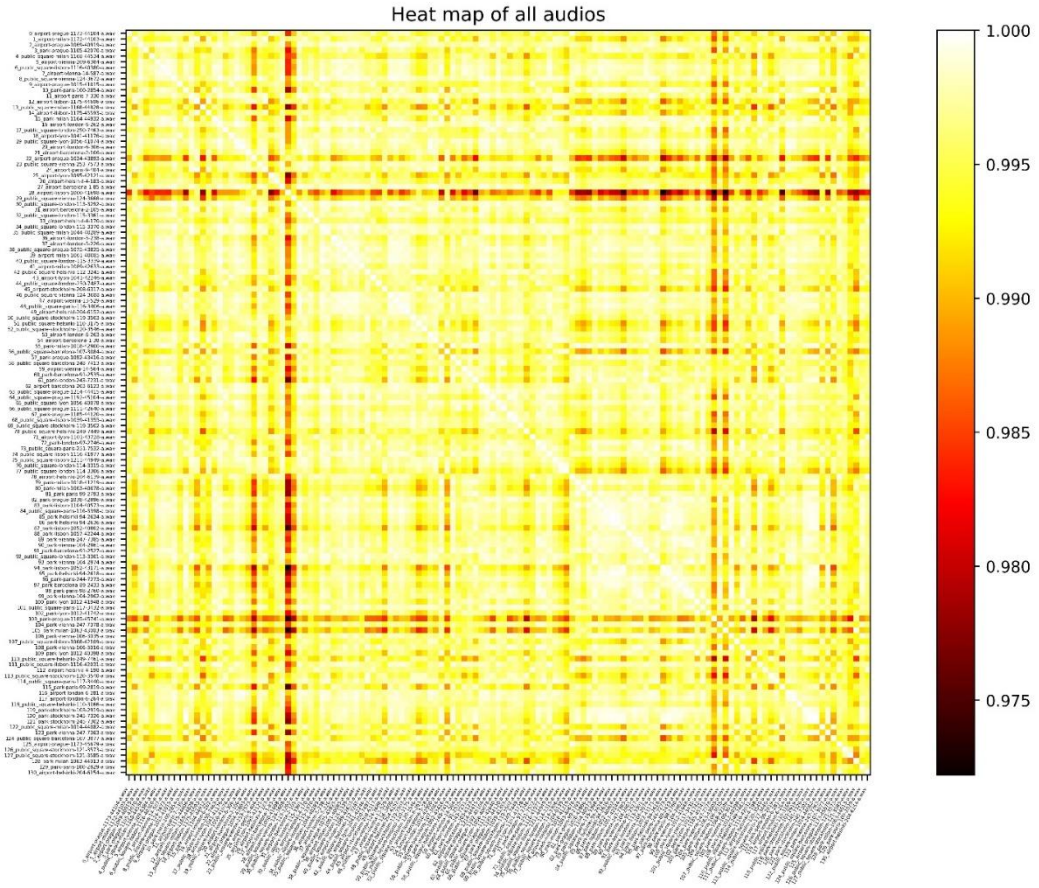

Figure 1. Heat map of all audios using cosine similarity

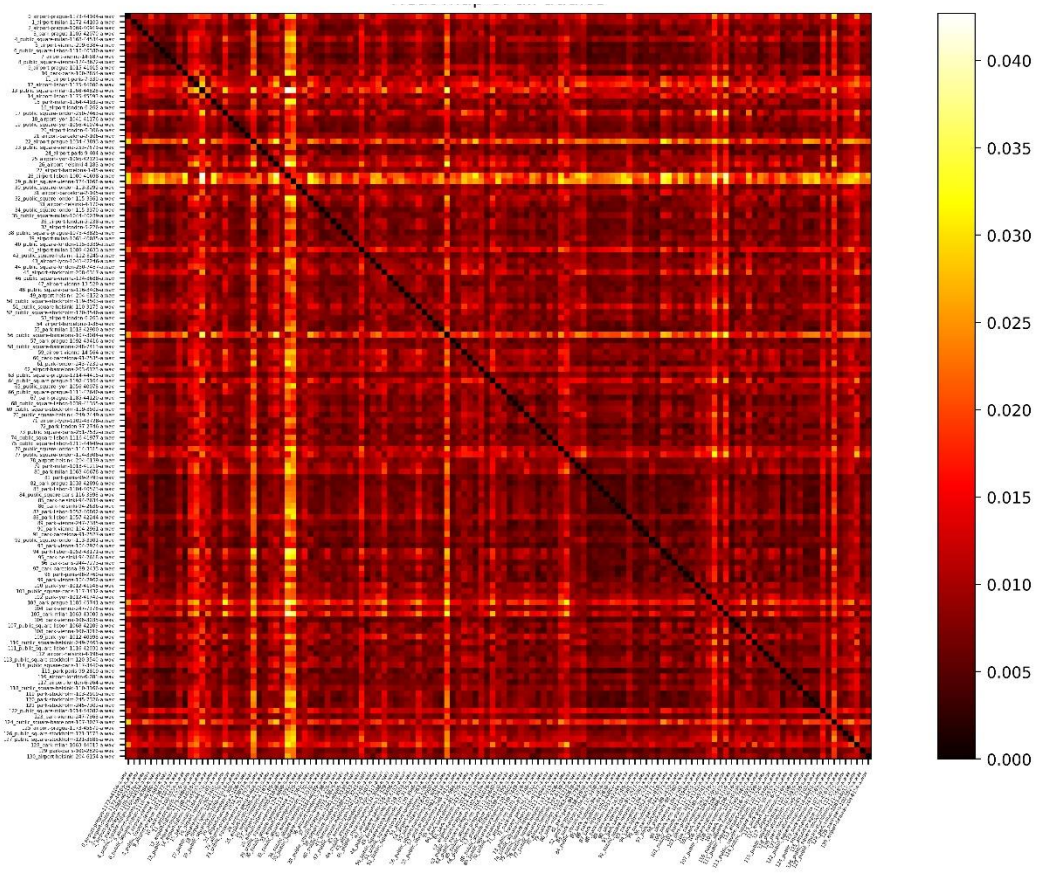

Figure 2. Heat map of all audios using DTW cosine distance

4. Conclusion

From my point of view, the project is interesting, despite the cost of time and effort. It provides practical knowledge and training to process real sound. I am aware of difficulties when analyzing these real data that it can cause confusing. To sum up, through this project, I have learnt the purposes of different analyzing methods, practiced using them and what information can be retrieved working with real audio.