

To explore the mechanism that allows music recognition apps  
to work.

Cephas Yeung

28/4/2025

---

## Contents

---

<b>1</b>	<b>Introduction</b>	<b>4</b>
<b>2</b>	<b>Methodology / Design Choice</b>	<b>5</b>
<b>3</b>	<b>Testing method</b>	<b>7</b>
<b>4</b>	<b>Discussion</b>	<b>8</b>
<b>5</b>	<b>Conclusion</b>	<b>9</b>
	<b>Bibliography</b>	<b>10</b>
<hr/>		
<b>A</b>	<b>Figures</b>	<b>12</b>
<b>B</b>	<b>Testing results</b>	<b>14</b>
<b>C</b>	<b>Screenshots</b>	<b>15</b>
<b>D</b>	<b>Meeting logs</b>	<b>16</b>

---

## Abstract

---

We recreated an algorithm used to recognise music.

The algorithm first fingerprints all the songs by create a spectrogram, generate peaks point inside it. To create an almost unique hash point where it's inserted inside a database.

When record songs through the interface, the algorithm repeats the process of fingerprinting, but instead compare hash value with matched songs

# CHAPTER 1

---

## Introduction

---

In modern society there are billions of songs created by different artists and different styles of music. Most of us might have already heard or have used the music recognition app Shazam<sup>®</sup>, or something similar on your mobile phone when you are at a restaurant or Café because you want to search for what song it was playing.

From using a very concrete source of a patent created by the original founder and Chief Scientist of Shazam<sup>®</sup><sup>1</sup>[8].

This source allows a great understanding of the background information of the program.

My aim for this research is to understand and replicate how modern day music recognition algorithms (explain in footnote) work.

During the creation process, it uses different resources to make the resource and also methods to test the success criteria set out by the algorithm.

In conclusion, the program is successful.

---

<sup>1</sup>Now part of Apple Inc<sup>®</sup>.

## CHAPTER 2

---

### Methodology / Design Choice

---


The artefact is separated in different pieces. As illustrated in the figures inside [Appendix A:Figure 1](#).


The software is run inside a container as the patent suggested that although not limited to any particular system, it is preferred for it to be run in a “Distributed system”<sup>1</sup>. But by containerize the program, it allows it to happen.

Next, the program is separated into two parts. The Web app part is for the user to interact with the algorithm graphically.

And the main algorithm are produced using python. Inside the program, the library of NumPy[5] to analysis the data.

The program is replicated as described in the patent by first recording the audio, this is done through the web app.

All the original audio is first sent to an algorithm to “fingerprint” the piece of music. The algorithm are consist of spectrogram produced from the SciPy signal library where it uses STFT<sup>2</sup>. [9] An array of data illustrated in the figure.  This arrays of data are being filtered through maximum filter in the SciPy library [9]. Which is illustrated in the figures.

The two arrays are compared to use to pick peak points. The peaks are used to construct the pairs shown here. 

The pairs of points are then hashed using the default hashing function<sup>3</sup> from python which generate an almost unique hash values in the program.

The hash point is than placed into a database alongside with its song name and where the point A’s real time. The song data (metadata) such as author and copyright detail are placed in a different database.

---

<sup>1</sup>Computer systems whose intercommunicating components are located on different networked computers[2]

<sup>2</sup>Short-time Fourier Transform

<sup>3</sup>Hashing function is a set of things to do to make an output that is always the same length from some other input data.[6]

When the user record the sample through the user interface created, the recording sample is sent to the server. It is processed where it applies the same algorithm from above. First it construct a spectrogram, create and compare it with a filtered spectrogram, peak point are picked and hash values are created, the hash value are then searched inside the ‘points‘ table.

The list of successfully matched song are than going through the process of ”scoring”. The score calculated by constructing a histogram with the point A’s real time value. The score is assigned to each song with the highest value in each histogram. This is because the peak point of some songs are the same for a completely different song. And this includes background noise where the algorithm could mistake it to be a different song. Illustrated in the `figure`.

[10, 7, 11]

## CHAPTER 3

---

### Testing method

---

The program is tested using the method described in this paper.[\[11\]](#) We used most of the songs in fingerprinted collection because that reduce the likelihood of a flawed test. We mainly used classical music because there is lower chance for creating copyright issues.

The test have increasing level of difficulty:

- Test 1: original song file (unmodified digital file)
- Test 2: original song file but shifted
- Test 3: original song file but with controlled level of noise. (E.g.: a sine wave)
- Test 4: original song file but played over the air and recorded.
- Test 5: original song but perform by different people.
- Test 6: transposed song

The test results is inside [B.1](#). The program overall worked as expected but test 6 doesn't work because this algorithm is not designed for this.

## CHAPTER 4

---

### Discussion

---

As discussed in [the test conducted](#), this artefact is not capable to detect a transposed song. Because this song only matches the exact frequency that the song produces.

One way that the problem can solve is to change the mechanism entirely to not rely purely on the raw frequency data, instead by using the MIDI file of the song which contains a series of message like note. [\[1\]](#) Then from it, create a database of melody. Which can recognise songs from it by recording the song and passing into the algorithm to process for results. This allows a wider error range for recognizing song such as made from "humming". [\[4\]](#) However, due to the time constraint of the project and we are trying to produce a replica of Shazam<sup>®</sup>. This algorithm might be much more costly because it requires the raw MIDI file or a construction of the MIDI file first, similarly performance might be poor from using MIDI file database.[\[11\]](#)



## CHAPTER 5

---

Conclusion

---

---

## Bibliography

---

- [1] amandaghassaei. *What Is MIDI?* Instructables. URL: <https://www.instructables.com/What-is-MIDI/> (visited on 01/31/2025).
- [2] *Distributed computing*. In: *Wikipedia*. Page Version ID: 1266361966. Dec. 31, 2024. URL: [https://en.wikipedia.org/w/index.php?title=Distributed\\_computing&oldid=1266361966#cite\\_note-tanenbaum-1](https://en.wikipedia.org/w/index.php?title=Distributed_computing&oldid=1266361966#cite_note-tanenbaum-1) (visited on 01/23/2025).
- [3] FFmpeg developers. *ffmpeg: A complete, cross-platform solution to record, convert and stream audio and video*. Version n7.1. URL: <https://www.ffmpeg.org/>.
- [4] Asif Ghias et al. “Query by humming: musical information retrieval in an audio database”. In: *Proceedings of the third ACM international conference on Multimedia - MULTIMEDIA '95*. the third ACM international conference. San Francisco, California, United States: ACM Press, 1995, pp. 231–236. ISBN: 978-0-89791-751-3. DOI: [10.1145/217279.215273](https://doi.org/10.1145/217279.215273). URL: <http://portal.acm.org/citation.cfm?doid=217279.215273> (visited on 01/31/2025).
- [5] Charles R. Harris et al. “Array programming with NumPy”. In: *Nature* 585.7825 (Sept. 17, 2020), pp. 357–362. ISSN: 0028-0836, 1476-4687. DOI: [10.1038/s41586-020-2649-2](https://doi.org/10.1038/s41586-020-2649-2). URL: <https://www.nature.com/articles/s41586-020-2649-2> (visited on 01/12/2025).
- [6] *Hash function*. In: *Simple English Wikipedia, the free encyclopedia*. Page Version ID: 9718140. Aug. 22, 2024. URL: [https://simple.wikipedia.org/w/index.php?title=Hash\\_function&oldid=9718140](https://simple.wikipedia.org/w/index.php?title=Hash_function&oldid=9718140) (visited on 01/30/2025).
- [7] Cameron MacLeod. *abracadabra: How does Shazam work?* - Cameron MacLeod. URL: <https://www.cameronmacleod.com/blog/how-does-shazam-work> (visited on 10/28/2024).
- [8] Danielle Newnham. *INTERVIEW: Dr Avery Wang, Principal Research Scientist at Apple and Co-Founder and Chief Scientist...* Medium. Feb. 2, 2023. URL: <https://daniellenewnham.medium.com/fostering-innovation-and-choosing-the-right-founding-team-6287f7eed9f> (visited on 02/02/2025).

- 
- [9] Pauli Virtanen et al. “SciPy 1.0: fundamental algorithms for scientific computing in Python”. In: *Nat Methods* 17.3 (Mar. 2020). Publisher: Nature Publishing Group, pp. 261–272. ISSN: 1548-7105. DOI: [10.1038/s41592-019-0686-2](https://doi.org/10.1038/s41592-019-0686-2). URL: <https://www.nature.com/articles/s41592-019-0686-2> (visited on 10/28/2024).
  - [10] Avery Li-Chun Wang and Julius O. Smith III. “Systems and methods for recognizing sound and music signals in high noise and distortion”. U.S. pat. 8386258B2. Shazam Investments Ltd. Feb. 26, 2013. URL: <https://patents.google.com/patent/US8386258B2/en?assignee=Shazam+Investments+Ltd&num=25> (visited on 12/25/2024).
  - [11] Cheng Yang. *Music Database Retrieval Based on Spectral Similarity*. Technical Report 2001-14. Backup Publisher: Stanford InfoLab. Stanford, 2001. URL: <http://ilpubs.stanford.edu:8090/489/>.

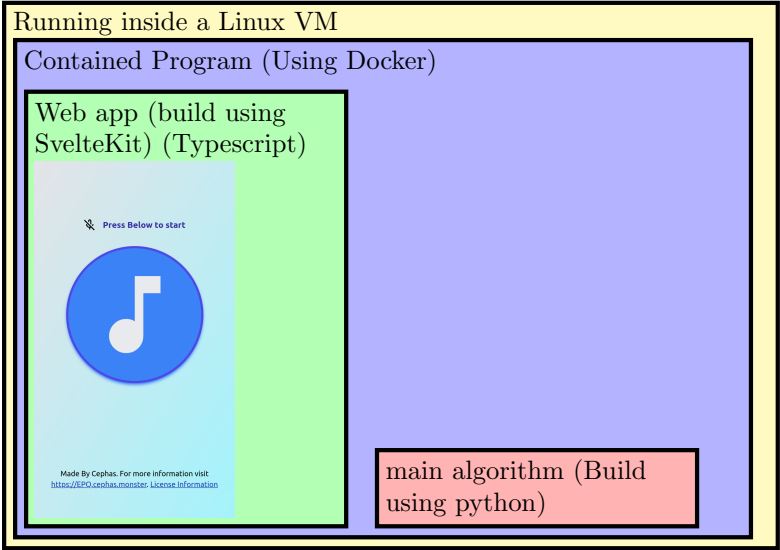


Figure A.1: Overall Architecture

0. User pressed start record button (🎵)

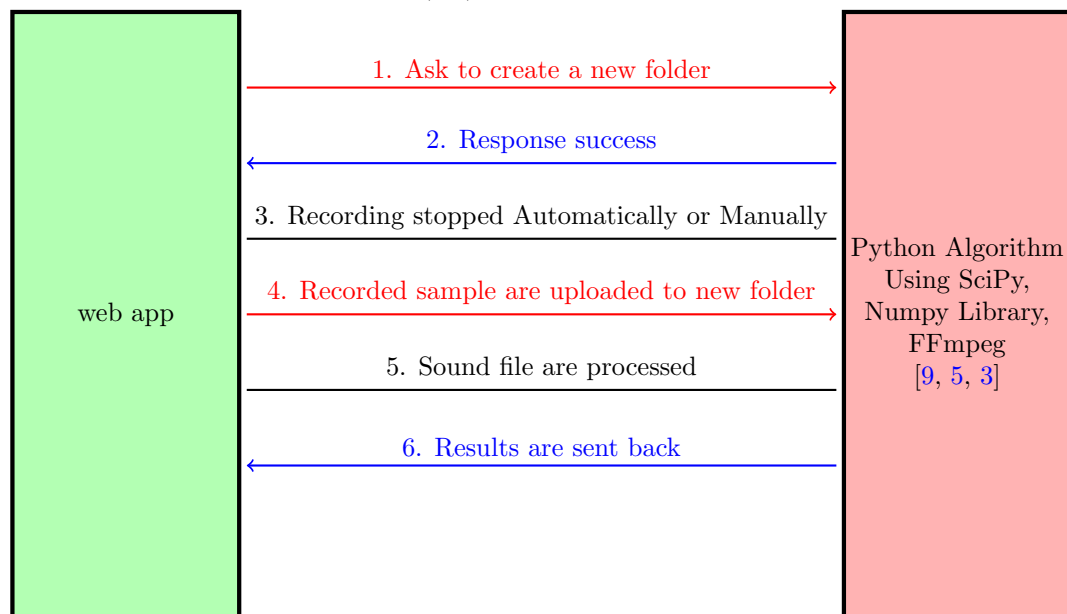


Figure A.2: Overall Architecture (app)

APPENDIX B

---

Testing results

---

filepath	passed	SUBTOTAL	filepath	passed	SUBTOTAL
test/test_1.py	46	46	test/test_2.py	46	46
TOTAL	46	46	TOTAL	46	46

Figure B.1: test result table

## APPENDIX C

---

### Screenshots

---

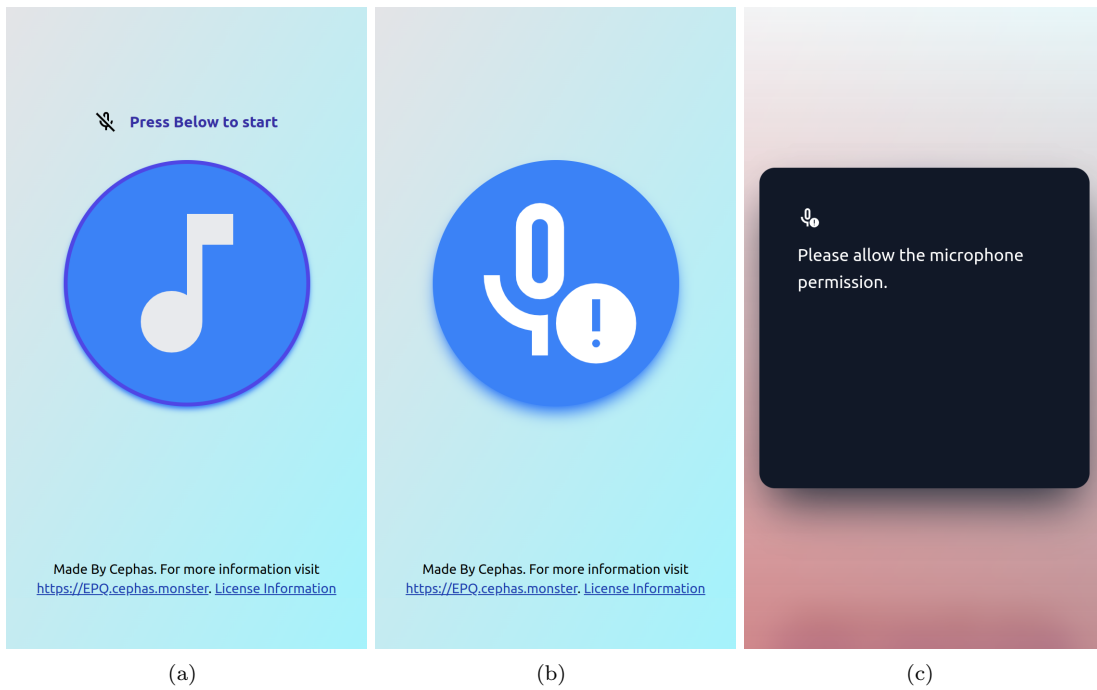


Figure C.1: (a) "Home", (b)(c) "Microphone permission are disabled"

All the figures are simulated to the size of an iPhone<sup>®</sup>SE display.

## APPENDIX D

---

### Meeting logs

---

#### Meeting with the supervisor at 25 November 2024

Summary:

Improve on the part A of the production logs, with more details on the titles described and what resources I will be using (such as the URLs should be provided), the timescale and what is involve.

As things I should be including to create a great project is the benefit that this might bring to the society and the impact.

And a plan to see what I should be sending or showing to the examiner in the future such as:

- Send product log
- Presentation of the artefact
- Evidence of the artefact
- Critical bibliography
- Place that in the Appendix and write a reference back to the appendix through the production log.

#### Meeting with a professional at 13 January 2025

Summary:

I explained the background details during the meeting.

I have explained that the program can only detect sounds from following test that I conducted such as:



- A test where there are no modifications to the original sound,
- A test where the song has cropped and reduced to a smaller size.
- A test where there is a constant noise generated by a program with a constant sine wave in the background.
- A test where the songs are mixed with a normal acceptable background noise such as coughing.

All of those above works as expected.

However the test that I did by recording when the sound is played on a music player on a different device doesn't work.

As an additional note, he has mentioned that the program is slow because of a reason which might be caused by the hash point not being an integer, as the hash value using integer is faster than using string as hash. (It was diagnosed with a slow database reading speed prior to the meeting).

The project has concluded as partially successful from the results produced from the searching algorithm.

**Summary of Improvement I need to make from the project.**

I should use a different hashing algorithm for the fingerprinting part of the program.

Round up the frequency, and the changes in time that are fed into the hashing algorithm.

**Additional note:**

The meetings are conducted with the consent of the professional. (With regards to the ethical issues)

The professional is my computer science teacher, as he has a decent qualification of a university degree.