

Automatic Synchronisation of Subtitle Track With Live Audio

Joshua Fenech

MLDM
Université de Jean Monnet
Saint-Étienne, France

M1 Masters Thesis 2018

Problem Motivation

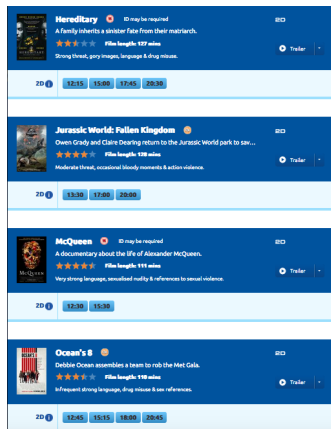


FIGURE – Full Film Showings 1 Day

Problem Motivation

Hereditary Ⓜ ID may be required
A family inherits a sinister fate from their matriarch.
★ ★ ★ ★ ★ Film length: 127 mins
Strong threat, gory images, language & drug misuse.

2D 12:15 15:00 17:45 20:30

Jurassic World: Fallen Kingdom Ⓜ
Owen Grady and Claire Dearing return to the Jurassic World park to save...
★ ★ ★ ★ ★ Film length: 128 mins
Moderate threat, occasional bloody moments & action violence.

2D 13:30 17:00 20:00

McQueen Ⓜ ID may be required
A documentary about the life of Alexander McQueen.
★ ★ ★ ★ ★ Film length: 111 mins
Very strong language, sexualised nudity & references to sexual violence.

2D 12:30 15:30

Ocean's 8 Ⓜ
Debbie Ocean assembles a team to rob the Met Gala.
★ ★ ★ ★ ★ Film length: 118 mins
Infrequent strong language, drug misuse & sex references.

2D 12:45 15:15 18:00 20:45

FIGURE – Full Film Showings 1 Day

ODEON
Covent Garden SW1Y Odeon

Hereditary subtitled (15)
Sun 01 July 15:00

Oceans 8 subtitled (12a)
Tue 03 July 18:00

FIGURE – Subtitled Film Showings 1 Week

2018-07-01

Patterns for SoS Reconfiguration

└ Introduction

└ Introduction

└ Problem Motivation

Problem Motivation



FIGURE – Full Film Showings 1 Day



FIGURE – Subtitled Film Showings 1 Week

Number of deaf people

Deaf people feeling excluded

Tourists

Aim

- Develop a method to watch subtitles on a phone
- Problem : Synchronising the subtitles to the film
- Therefore, must identify the time in the film based on audio signals

Prior Knowledge

```
gotS07E01.srt x
1 1
2 00:00:03,202 → 00:00:08,327
3 ♪ (PIANO PLAYING) ♪
4
5 2
6 00:00:13,452 → 00:00:14,493
7 ♪ (PIANO STOPS) ♪
8
9 3
10 00:00:14,827 → 00:00:16,827
11 -(EXPLOSIONS)
12 -(MAN YELLING)
13
14 4
15 00:00:17,452 → 00:00:19,493
16 ♪ (PIANO PLAYING) ♪
17
18 5
19 00:00:21,952 → 00:00:24,493
20 The war is over.
21 Winter has come.
22
23 6
24 00:00:24,785 → 00:00:26,202
25 JON SNOW: The war is not over.
26
27 7
28 00:00:26,202 → 00:00:28,493
29 The true enemy won't
30 wait out the storm.
31
```

Patterns for SoS Reconfiguration

└ Introduction

Introduction

- └ Prior Knowledge

[illegible]

Who here has pirated a film ?

Used subtitles ?

srt

General Method

- Record audio, compressed using MP3
- Split signal into frames of duration 25ms - consider signal constant over this period
- Extract Mel Frequency Cepstral Coefficients (MFCC's)
- Use MFCC's as predictive feature of whether speech is present in a frame or not
- Match these predictions to the truth array, defined by a subtitle file

Prior Knowledge

- How do you use subtitles on a laptop ?
- SubRip Subtitle file (.srt)

MP3 Compression

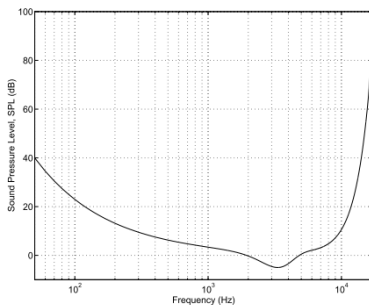


FIGURE – Frequency respons of human hearing

- Absolute threshold of hearing

MFCC Audio Features

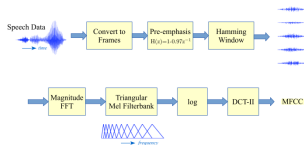


FIGURE – Sourced from

- Process based on psychoacoustics to represent features most important to human hearing
- Split audio file into small sections, consider features constant over this period of time
- Apply a series of transformations
- Reduce stuff

MFCC Audio Features

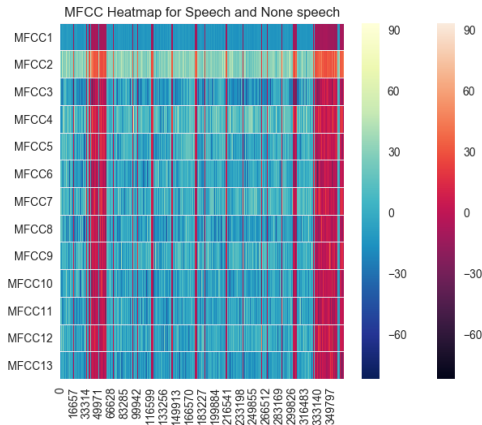


FIGURE – MFCC's Game of Thrones

Learner Architecture

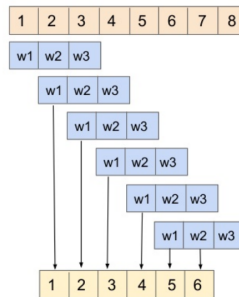
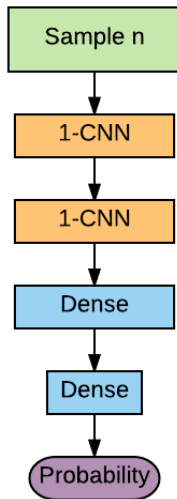


FIGURE – 1d convolutions, no padding

FIGURE – Model architecture

Results

- Learner trained on Game of Thrones episode
- Results on validation data suggested this was sufficient

Synchronisation

- Access to dataset granted incrementally as new audio is recorded
- Initially attempted to match a window of predicted probabilities with a similar array generated from srt
- Problem : Beginning of film often has no subtitle
- Solution : Continue recording data until speech is detected, and identify this as start of subtitle track