

# Automatic Synchronisation of Subtitle Track With Live Audio

Joshua Fenech

MLDM  
Université de Jean Monnet  
Saint-Étienne, France

M1 Masters Thesis 2018

# Problem Motivation

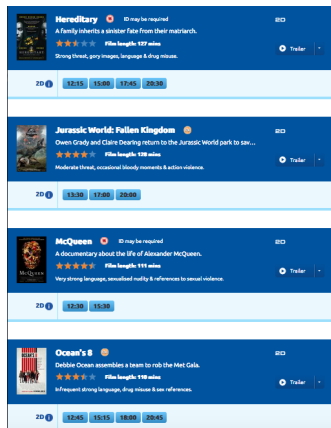


FIGURE – Full Film Showings 1 Day

# Problem Motivation

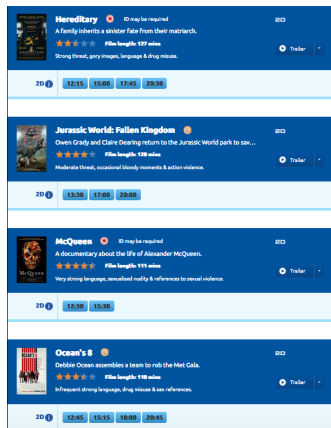


FIGURE – Full Film Showings 1 Day

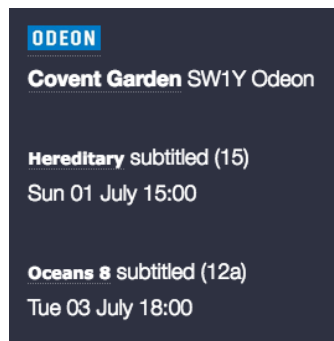


FIGURE – Subtitled Film Showings 1 Week

2018-06-28

# Patterns for SoS Reconfiguration

- └ Introduction

- └ Introduction

- └ Problem Motivation

Number of deaf people

Deaf people feeling excluded

Tourists

## Problem Motivation



FIGURE – Full Film Showings 1 Day



FIGURE – Subtitled Film Showings 1 Week

# 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
```

# Prior Knowledge

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

# MP3 Compression

- MPEG-III (mp3) Applies a load of compression stuff
- Drops frequencies outside of human hearing
- Other stuff

# MFCC Audio Features

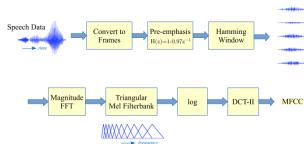


FIGURE – Sourced from

- 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

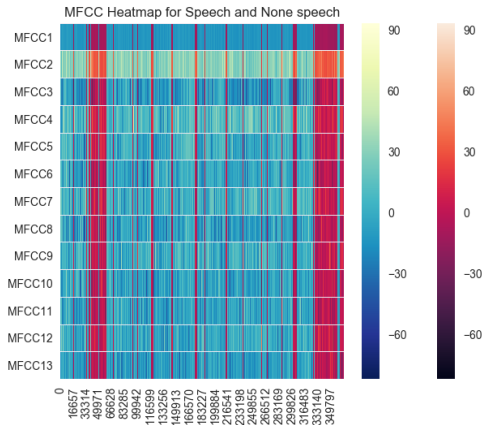


FIGURE – MFCC's Game of Thrones

# Learner Architecture

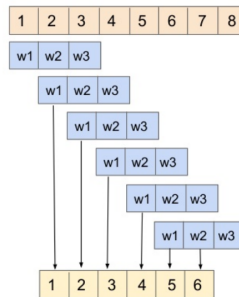
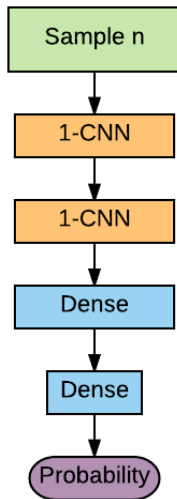


FIGURE – 1d convolutions, no padding

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