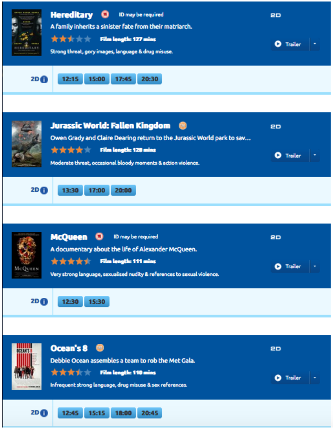
Automatic Sychronisation of Subtitle Track With Live Audio

Joshua Fenech

MLDM
Université de Jean Monnet
Saint-Étienne, France

M1 Masters Thesis 2018

Problem Motivation



Hereditary ID may be required — 2D

A family inherits a sinister fate from their matriarch.

Film length: 127 mins

Strong threat, gory images, language & drug misuse.

Trailer

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

Jurassic World: Fallen Kingdom — 2D

Owen Grady and Claire Dearing return to the Jurassic World park to sav...

Film length: 128 mins

Moderate threat, occasional bloody moments & action violence.

Trailer

2D 13:30 17:00 20:00

McQueen ID may be required — 2D

A documentary about the life of Alexander McQueen.

Film length: 111 mins

Very strong language, sexualised nudity & references to sexual violence.

Trailer

2D 12:30 15:30

Ocean's 8 — 2D

Debbie Ocean assembles a team to rob the Met Gala.

Film length: 110 mins

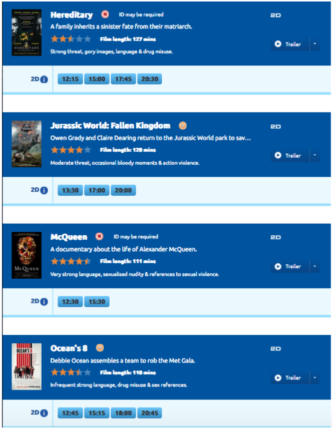
Infrequent strong language, drug misuse & sex references.

Trailer

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

FIGURE – Full FIlm Showings 1 Day

Problem Motivation



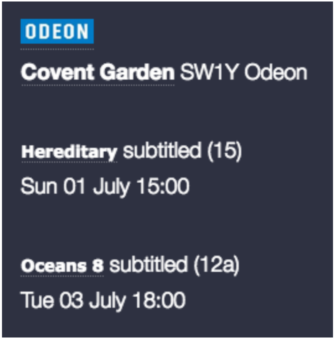
Hereditary ID may be required 2D
A family inherits a sinister fate from their matriarch.
Film length: 127 mins
Strong threat, gory images, language & drug misuse.
Trailer
2D 12:15 15:00 17:45 20:30

Jurassic World: Fallen Kingdom 2D
Owen Grady and Claire Dearing return to the Jurassic World park to sav...
Film length: 128 mins
Moderate threat, occasional bloody moments & action violence.
Trailer
2D 13:30 17:00 20:00

McQueen ID may be required 2D
A documentary about the life of Alexander McQueen.
Film length: 111 mins
Very strong language, sexualised nudity & references to sexual violence.
Trailer
2D 12:30 15:30

Ocean's 8 2D
Debbie Ocean assembles a team to rob the Met Gala.
Film length: 110 mins
Infrequent strong language, drug misuse & sex references.
Trailer
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

Problem Motivation

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

Number of deaf people

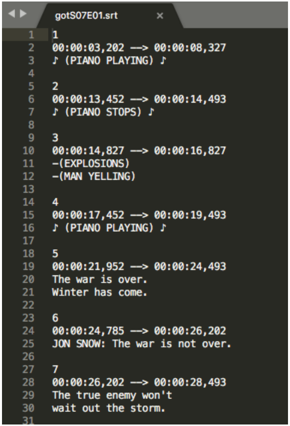
Deaf people feeling excluded

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

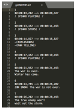
Prior Knowledge



```
1
00:00:03,202 --> 00:00:08,327
♪ (PIANO PLAYING) ♪

2
00:00:13,452 --> 00:00:14,493
♪ (PIANO STOPS) ♪

3
00:00:14,827 --> 00:00:16,827
-(EXPLOSIONS)
-(MAN YELLING)

4
00:00:17,452 --> 00:00:19,493
♪ (PIANO PLAYING) ♪

5
00:00:21,952 --> 00:00:24,493
The war is over.
Winter has come.

6
00:00:24,785 --> 00:00:26,202
JON SNOW: The war is not over.

7
00:00:26,202 --> 00:00:28,493
The true enemy won't
wait out the storm.
```

FIGURE – SubRip .srt File

Prior Knowledge



```
gotS07E01.srt
1
00:00:03,202 --> 00:00:08,327
♪ (PIANO PLAYING) ♪

2
00:00:13,452 --> 00:00:14,493
♪ (PIANO STOPS) ♪

3
00:00:14,827 --> 00:00:16,827
-(EXPLOSIONS)
-(MAN YELLING)

4
00:00:17,452 --> 00:00:19,493
♪ (PIANO PLAYING) ♪

5
00:00:21,952 --> 00:00:24,493
The war is over.
Winter has come.

6
00:00:24,785 --> 00:00:26,202
JON SNOW: The war is not over.

7
00:00:26,202 --> 00:00:28,493
The true enemy won't
wait out the storm.
```

FIGURE – SubRip .srt File

Who here has pirated a film ?

Used subtitles ?

srt

Subrip files contain list of entries indicating start time, stop time and text to be displayed

General Method

- Record audio, compressed using MP3
- Split signal into frames of duration 25ms - consider signal constant over this period
- Take frames every 10ms, so frames overlap
- Extract Mel Frequency Cepstral Coefficients (MFCC's) from each frame
- 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 srt file

Prior Knowledge

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

MP3 Compression

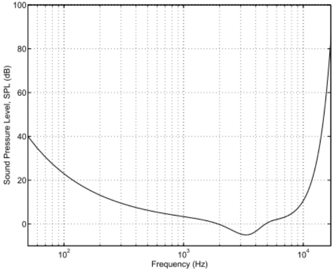


FIGURE – Frequency response of human hearing [3]. Curve indicates amplitude required to detect tone at a given frequency.

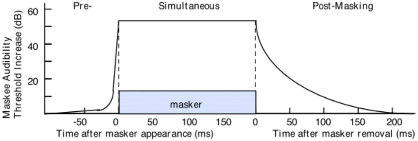


FIGURE – Frequencies masked by more prevalent frequencies[3].

- Create array of length appropriate to video with each entry corresponding to a frame

- Compare entries of srt to these start/stop times and ascribe a 1 if subtitles are present

Algorithm 1 pb_array_fill

```
procedure
    $i \leftarrow 0$
    $j \leftarrow 0$
    $m \leftarrow pb\_array\_length$
    $n \leftarrow subs\_array\_length$
    while True do
        if $i > m$ then
        if $j > n$ then
        if $pb\_array[i]\ start\ time \geq subs[j]\ start\ time$ then
            if $pb\_array[i]\ end\ time < subs[j]\ end\ time$ then
                $pb\_array[i] \leftarrow 1$
                $i \leftarrow i + 1$
        else
            $j \leftarrow j + 1$
```

MFCC Audio Features

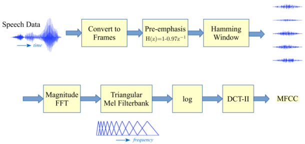


FIGURE – Steps of MFCC[2]

- 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

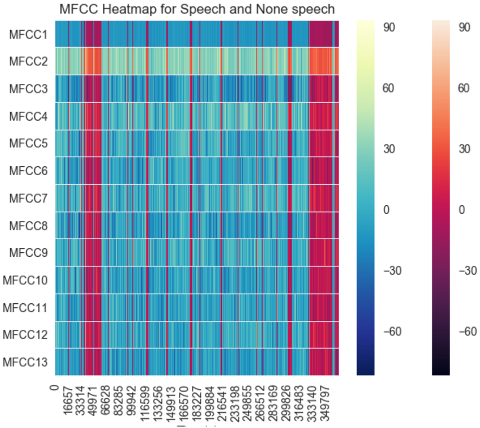


FIGURE – MFCC's Game of Thrones

Learner Architecture

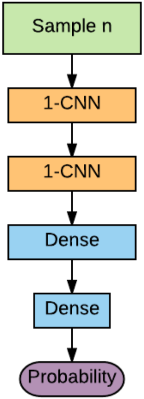


FIGURE – Model architecture[4]

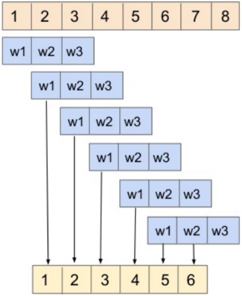


FIGURE – 1d convolutions, no padding [1]

Results

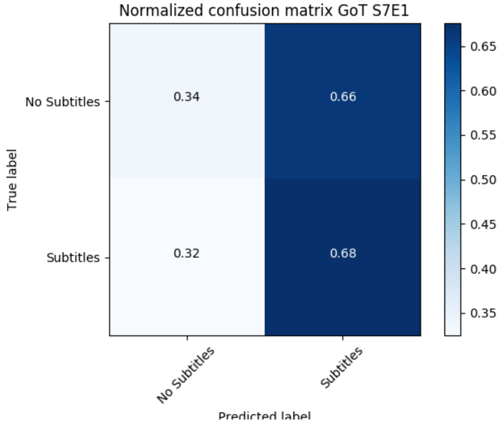


FIGURE – Confusion Matrix Game of Thrones

Results

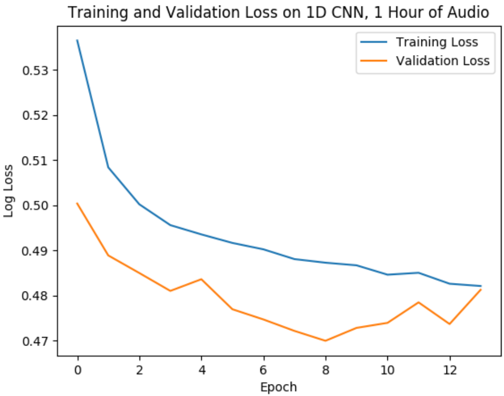


FIGURE – Training error

Results

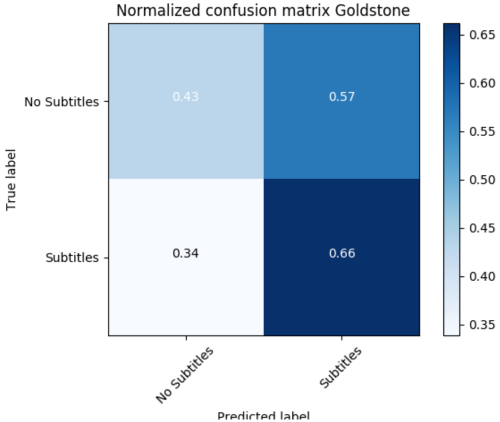


FIGURE – Test Time

Results

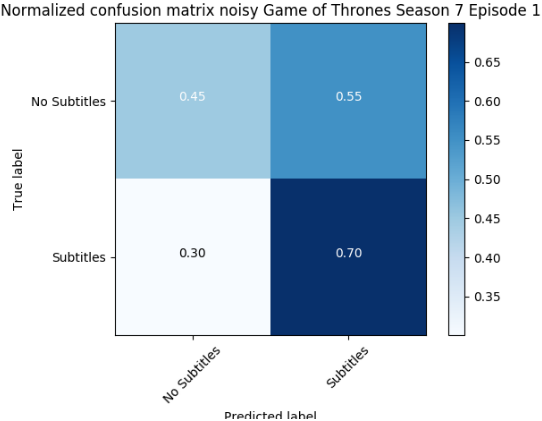
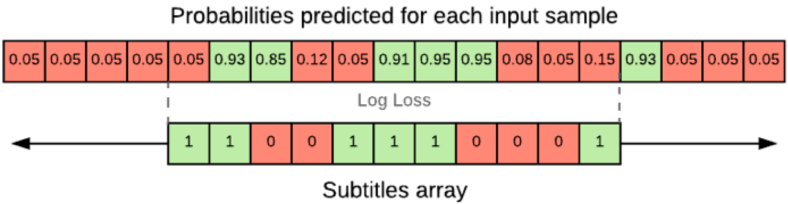


FIGURE – Test Time

Array Matching



Probabilities predicted for each input sample

0.05	0.05	0.05	0.05	0.05	0.93	0.85	0.12	0.05	0.91	0.95	0.95	0.08	0.05	0.15	0.93	0.05	0.05	0.05

1	1	0	0	1	1	1	0	0	0	1

Subtitles array

FIGURE – Match predictions with truth array using log loss [4]

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

Future Work

- Improve accuracy
- Remove nonspeech subtitles
- More efficient search algorithm
- Implement multithreading so that audio can be recorded and features extracted concurrently
- Alternative languages