Investigating the Viability of Automated Quality Control for Subtitles

Michael Armstrong

Associate staff - University of Dundee
& member of the UK Subtitling Audiences Network

SMPTE Media Broadcast Technology Conference Europe
13/5/2026

My Background

Between 2012 and 2016 I led a small team at BBC R&D investigating live subtitle quality.

www.bbc.co.uk/rd/projects/live-subtitle-quality

After taking redundance in 2024 I became aware that some of the improvements we had made to subtitles were no longer effective.

In order to demonstrate the problem I needed to make quality measurements over long periods of time.

Definitions

This presentation is only about television subtitles in the UK.

Also known as subtitles for deaf and hard of hearing (SDH) and closed captions (CC).

They are subtitles in the **same** language as the speech with additional information such as sound effects.

Subtitles for translation and translation SDH are entirely different.

How are TV subtitles made?

- Manual typing – QUERTY keyboard (prerecorded only)
- Stenography – typing in shorthand (live only)
- Respeaking – single-voice speech-to-text (live and prerecorded)
- Formatted from autocue & scripts – (live and prerecorded)
- Automatic Speech Recognition – direct speech to text (live only)

Key problems with subtitle quality

- **Timing** – live subtitles usually arrive several seconds behind the speech and occasionally a few seconds early.

- **Word loss** - Deaf and hard-of-hearing people are consistent in expressing their preference for verbatim subtitles. The problem is that the omission and substitution of words disrupt the viewer's attention.

- **Loss of subtitles** – e.g. technical failures

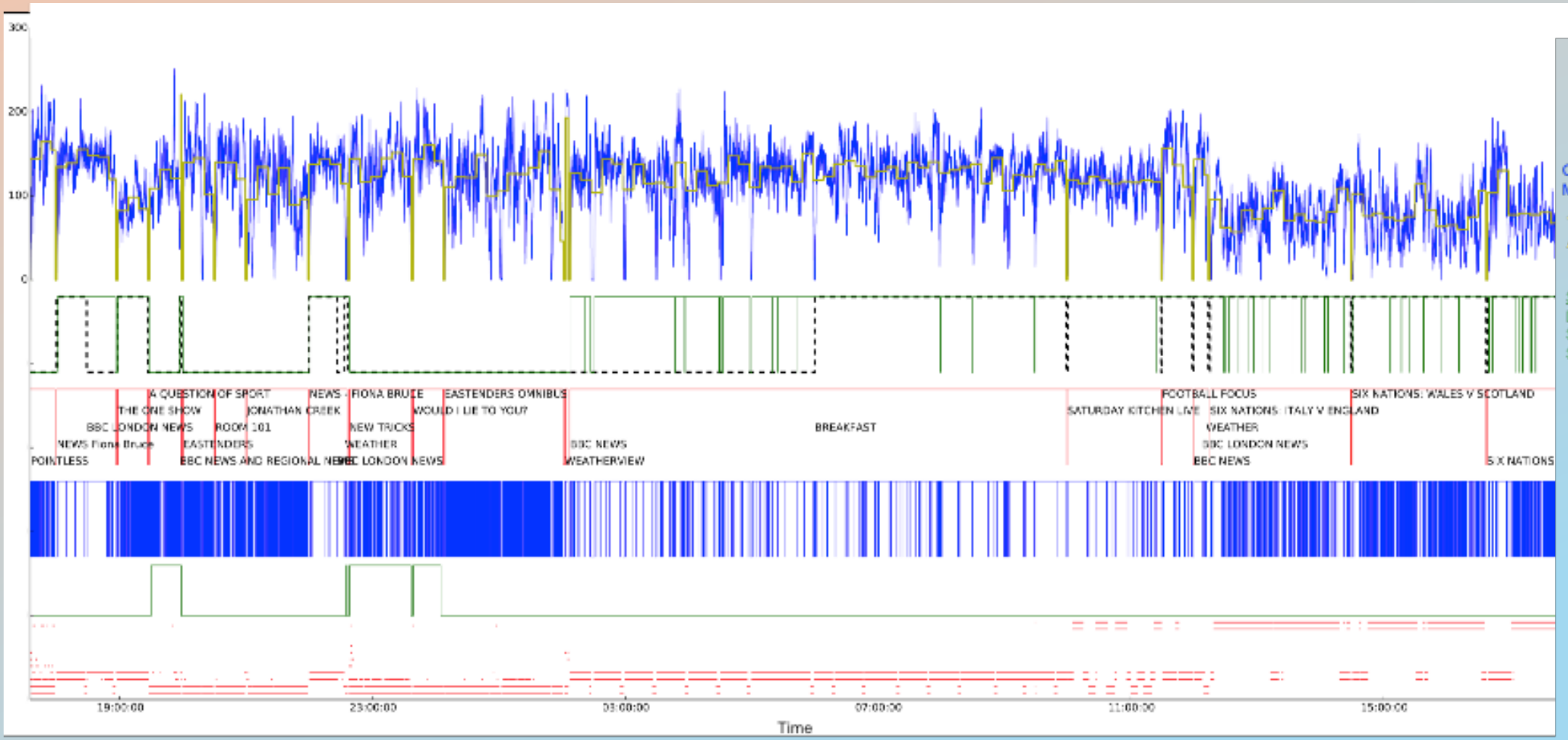
Manual subtitle quality monitoring

Manual surveys are expensive and time consuming examining short samples and usually focusing on word errors.

In one by *Ofcom* and the *University of Roehampton* in 2014/15. 10-min clips were tested at 6-month intervals over two years.

The extremely sparce sampling, questionable methodology and the high costs meant the exercise caused a severe backlash.

24/7 measurement - BBC R&D 2014



Adding speech to text

Modern speech to text engines are not 100% accurate and will interpret some speech content differently from a human subtitler.

However, they can give accurate word timings and a reasonably accurate word count.

This project uses a modified version of *Open AI's Whisper* called *whisper_timestamped*.

Automating subtitle quality measurement

The errors in TV subtitles are likely to be different from those in a speech-to-text transcript.

By comparing the two, the difference exposes problems in the subtitles, in particular:

- **Timing** – how late or early are the subtitles?

&

- **Word count** – how many words are missing from the subtitles?

The workflow

This project is written in python 3 and runs off-line.

The test material is transport stream recordings from UK *Freesat* services, which carry Teletext subtitles.

The main audio track and Teletext subtitles are extracted using *ffmpeg* to a .wav file for the audio and subtitles as a .srt file.

The audio is then passed to *Whisper* to produce a transcript.

Subtitles are not structured data

Subtitles describe how text should be displayed on a screen.

They contain repetition and non-speech elements which need to be separated from the speech content.

The subtitle file is converted into a structured data format where repeats are removed and non-speech components are stored separately. This process is not 100% reliable.

Alignment

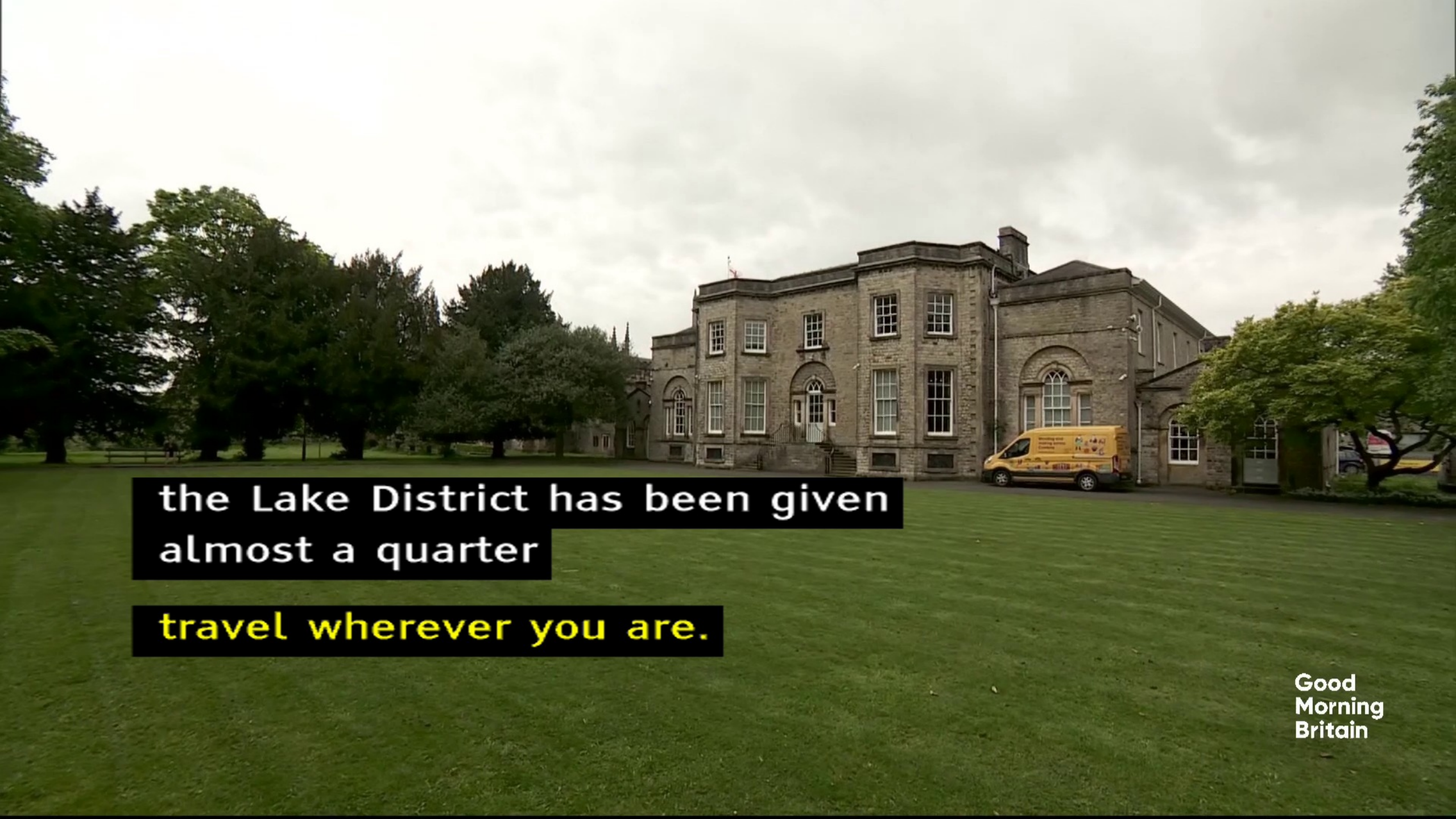
To measure timing, the transcript output by the speech-to-text engine first has to be aligned to the television subtitles.

- This is straightforward with high quality, pre-prepared subtitles and a clear speech soundtrack.

- However, as the subtitle and audio quality decline, the difficulty of aligning the transcript to the subtitles increases.

Things that make alignment difficult - 1

- The timing in the subtitles and transcript may not match.
- The subtitles may omit many of the spoken words.
- Word errors in both the subtitles and the transcript.
- Spelling differences between the subtitles and the transcript.
- Compound words and contractions vs separate words.
- Numerals vs number words.
- The words in the subtitles can be in the wrong order.
- Long sections of subtitles can be repeated.
- Subtitles can get stuck.


the Lake District has been given
almost a quarter
travel wherever you are.
Good
Morning
Britain

Things that make alignment difficult - 2

- Live subtitles can contain corrections – in varying formats.
- Non-speech utterances, especially in children's TV.
- Speech or singing containing lots of repetition.

Also, the software needs to cope with...

- Channels with no subtitle stream.
- Programmes with no subtitles.
- Programmes that contain subtitles but no speech.
- Programmes broadcast with the wrong subtitles.

Alignment process

- The software first looks in 4 minute sections for the longest continuous string of words (n-grams) that occurs only once in both the subtitles and transcript and then the next longest and so on down to around 20 word strings.

- It then does the same in the remaining unmatched gaps.

- A final pass checks for differences in spelling, numerals, compound words and contractions, ignoring the most common 20 words to avoid false alignments.

Calibration

The accuracy of the system was confirmed by the measurements it gives on known high quality subtitles, i.e. pre-recorded, factual programmes with a clear speech content.

- The word count differences are less than ±1%.
- The timing measurements are within ±1 second.
- The software manages to align over 97% of the words.

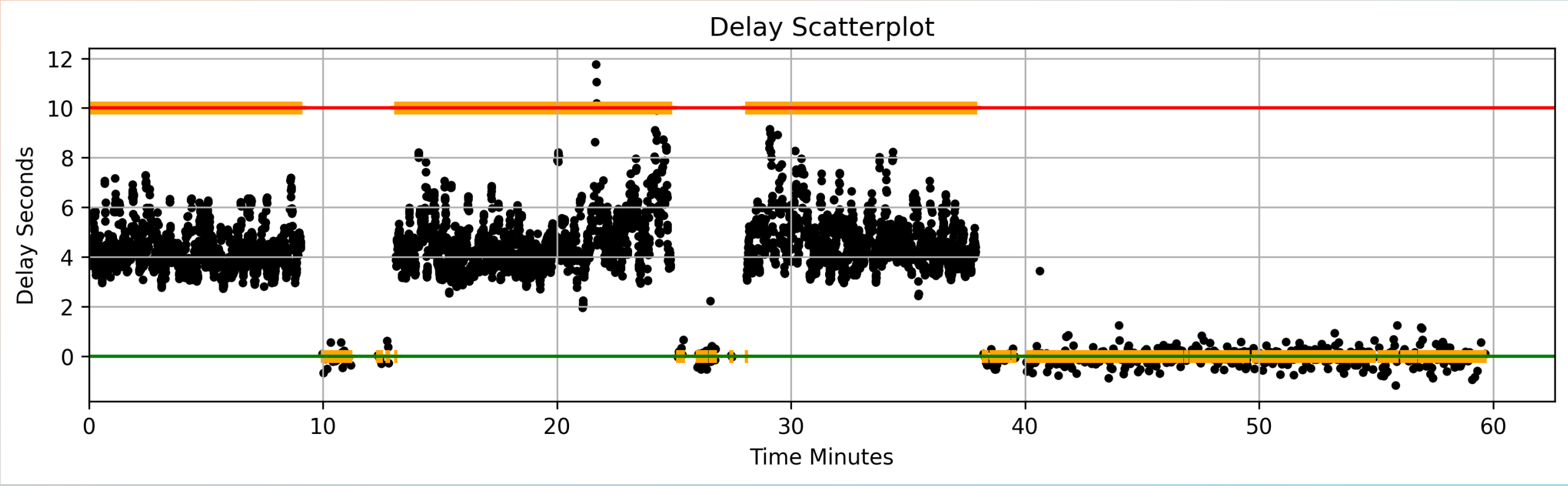
Displaying the results

The results are plotted against a time-line in a series of graphs.

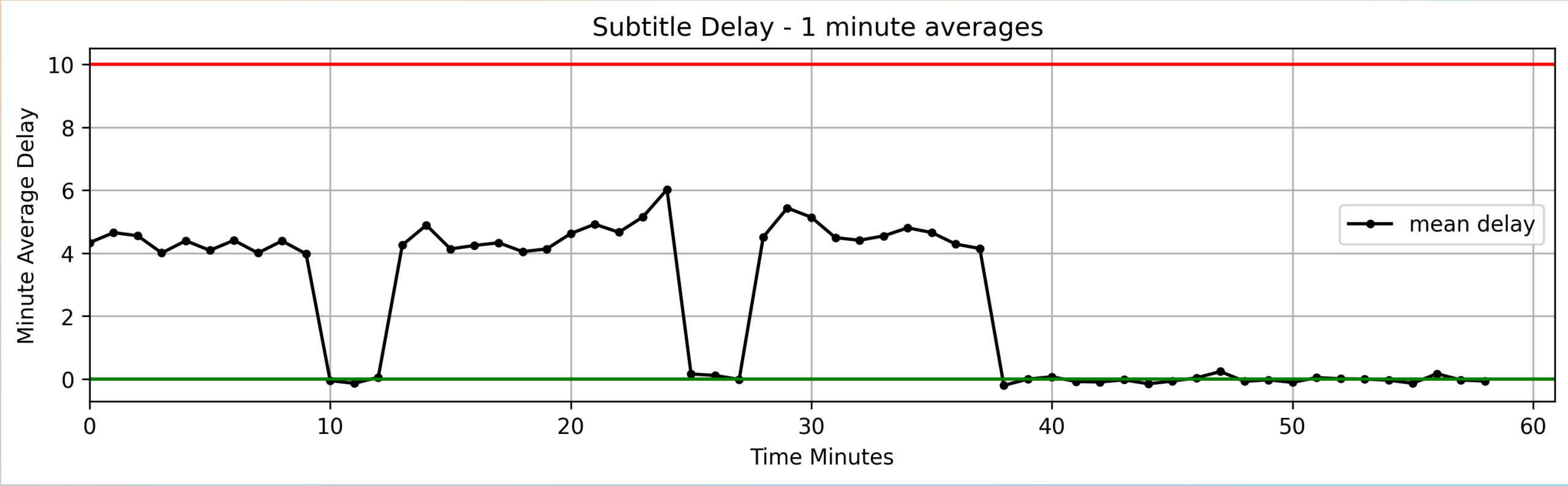
Subtitles over 10 seconds late are effectively useless, indicated by a red line on the plot with a green line for zero delay.

Orange markers on the red line indicate snake/word-at-a-time subtitles and orange markers on the green line indicate block subtitles.

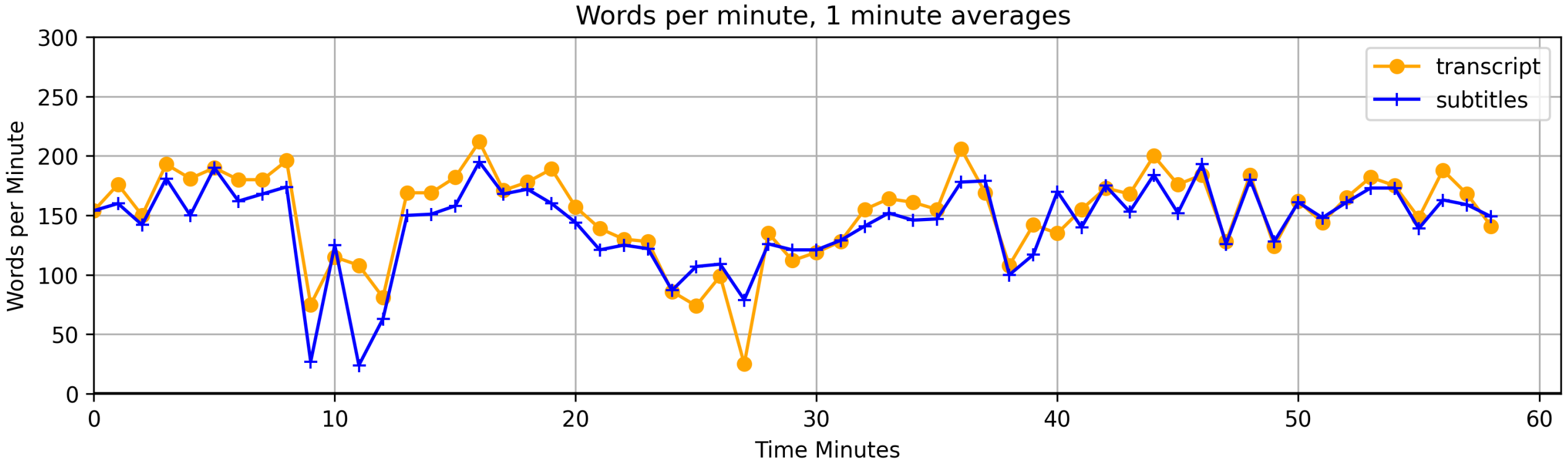
Timing - scatter plot



Timing - one minute-average



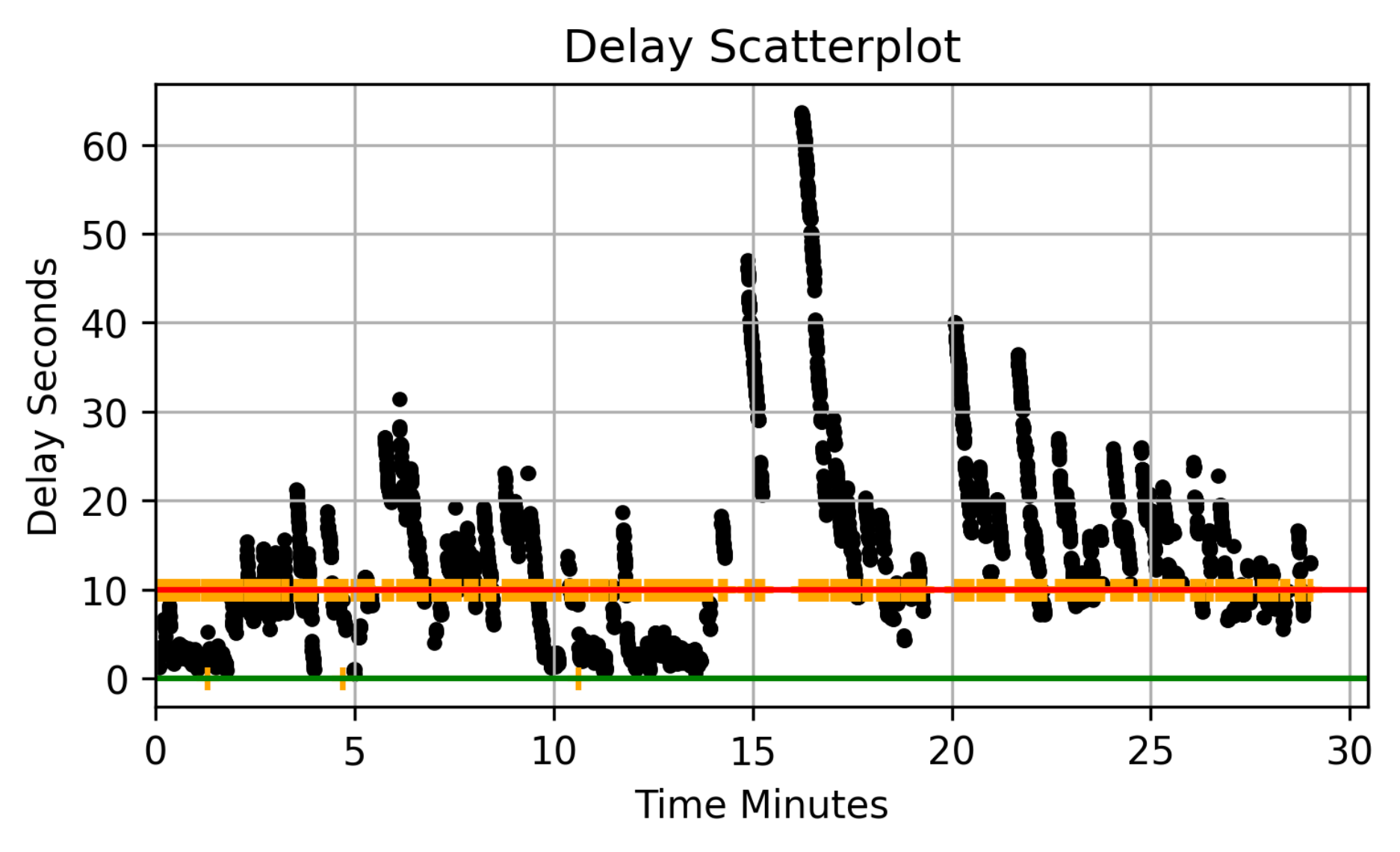
Words in each minute



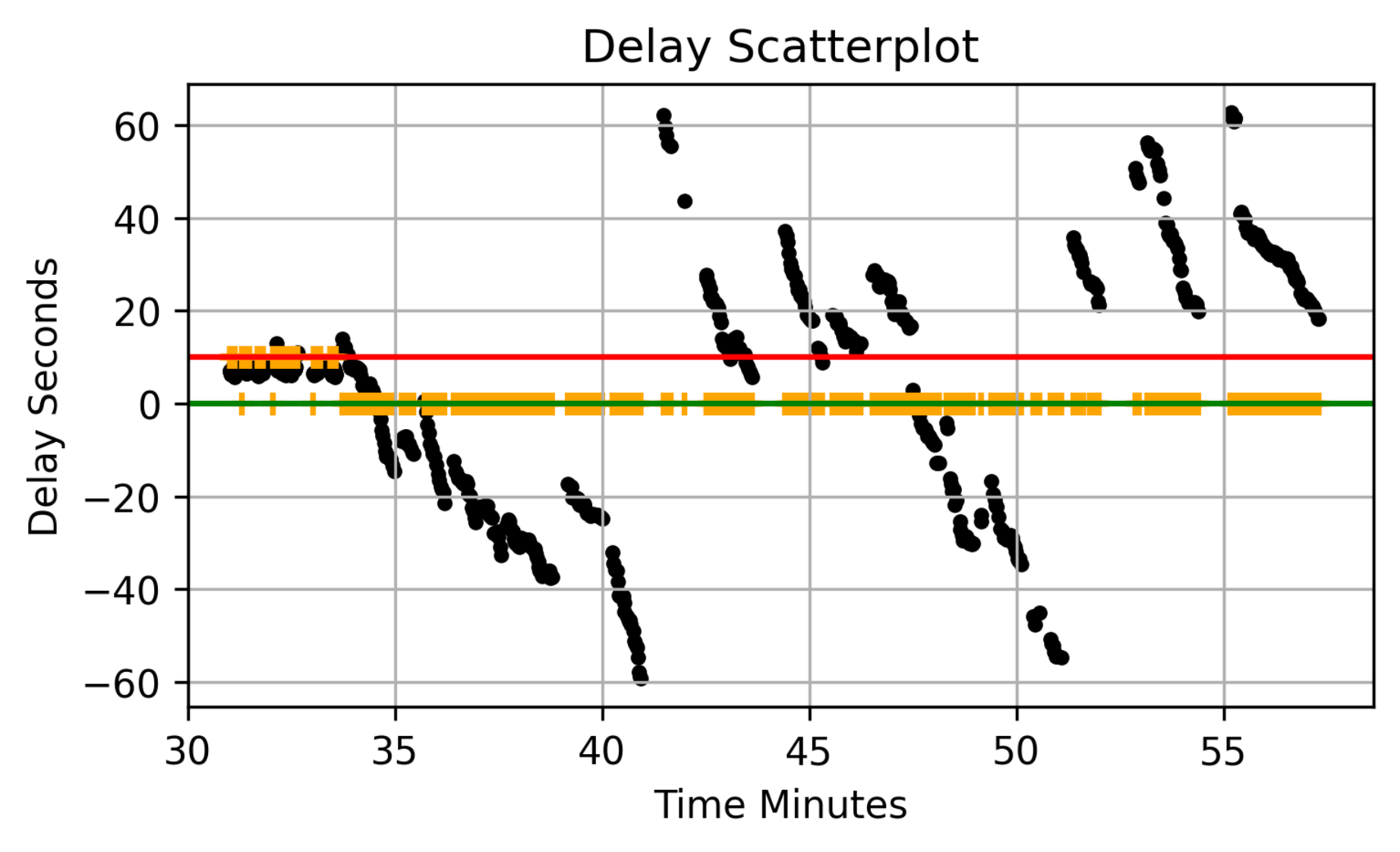
Quality failings

- Some channels regularly broadcast subtitles over 10 seconds late, sometimes many times that.
- There also are examples of subtitles being broadcast early.
- And examples of subtitles omitting over half the spoken words.

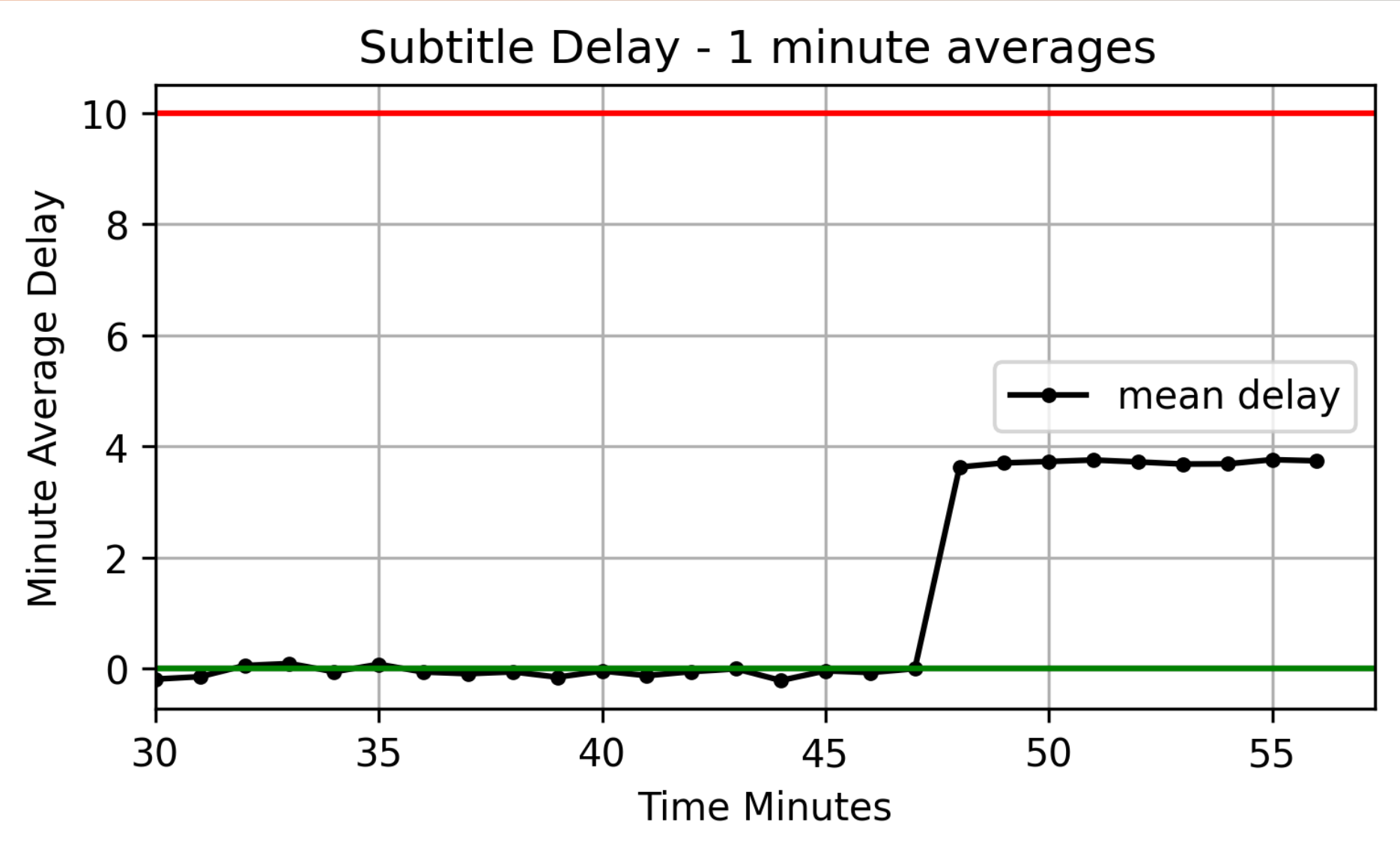
Problems with timing – live (snake)



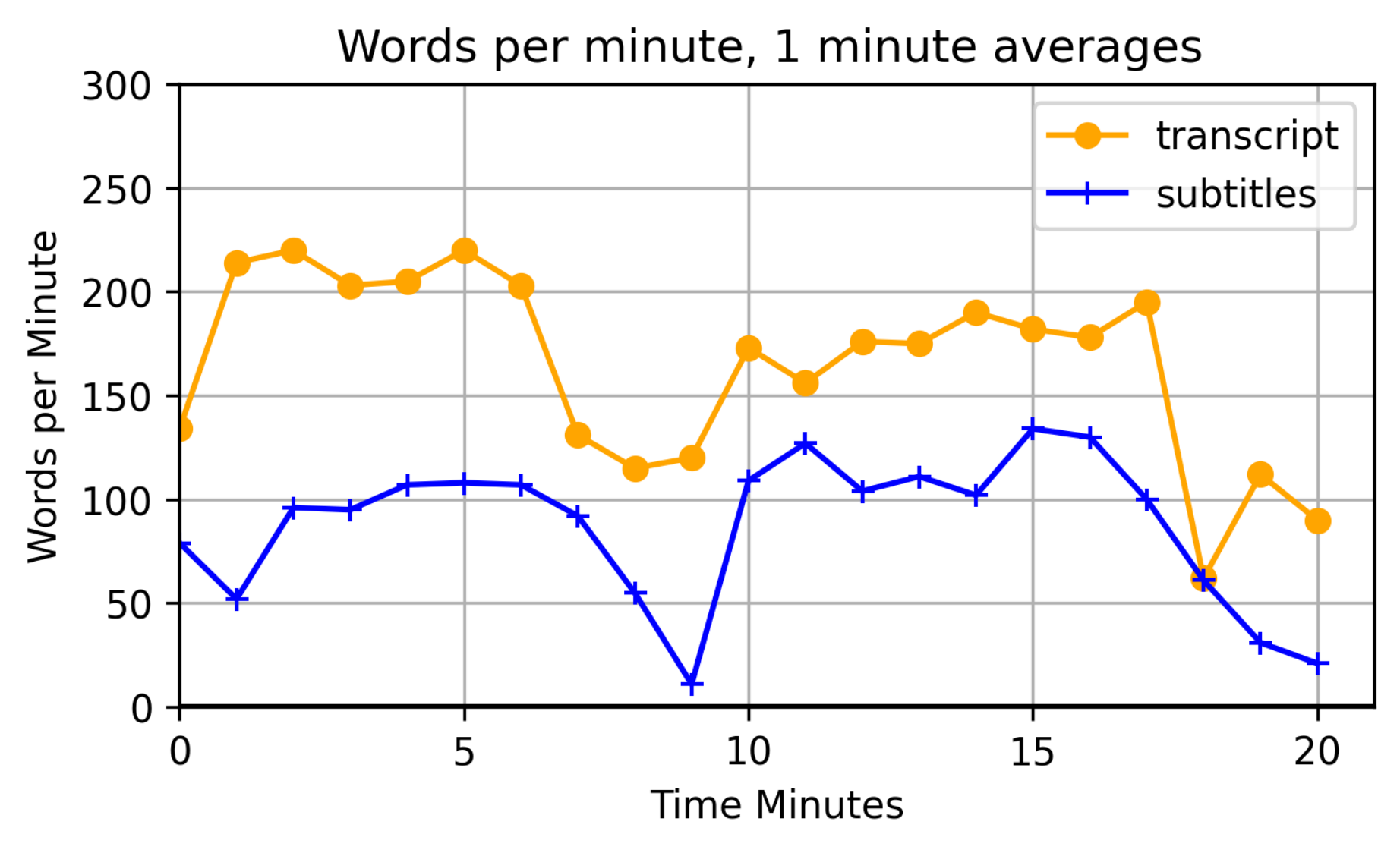
Problems with timing – live (block)



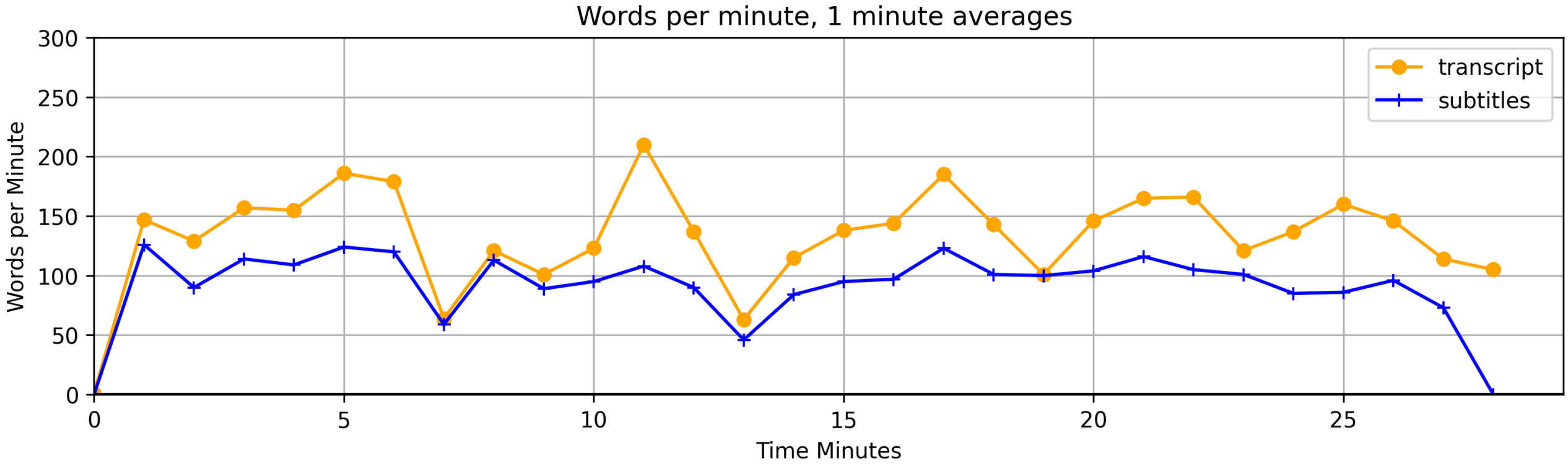
Problems with timing – pre-recorded



Problems with word loss - live



Problems with word loss – archive (1988)

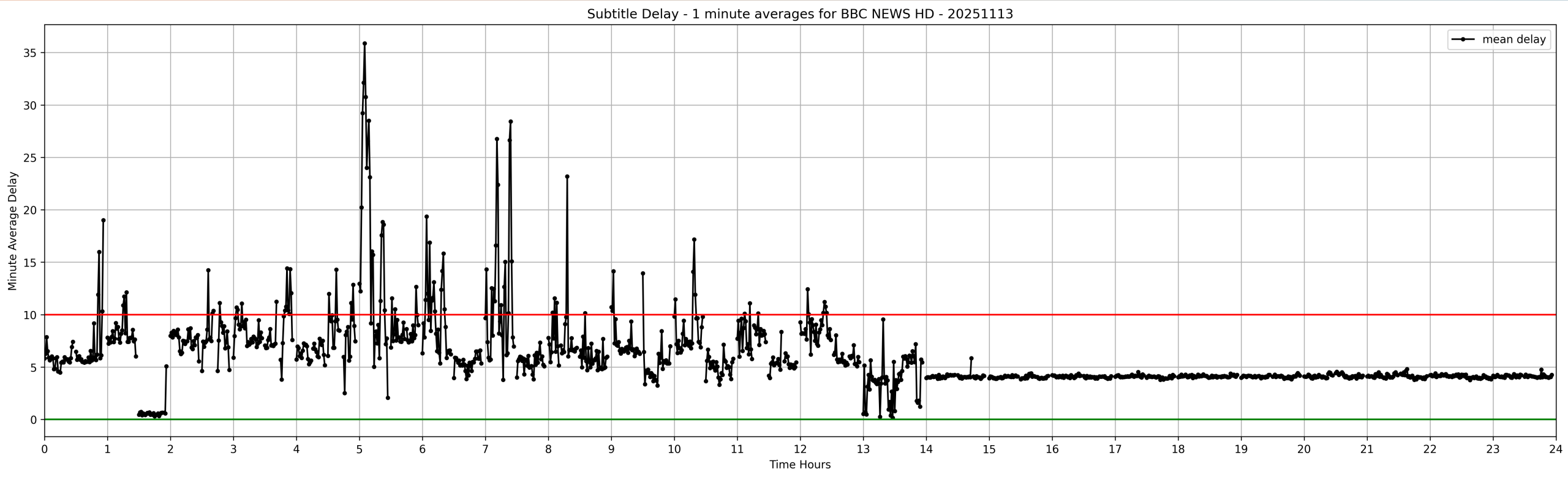


24 hour plots

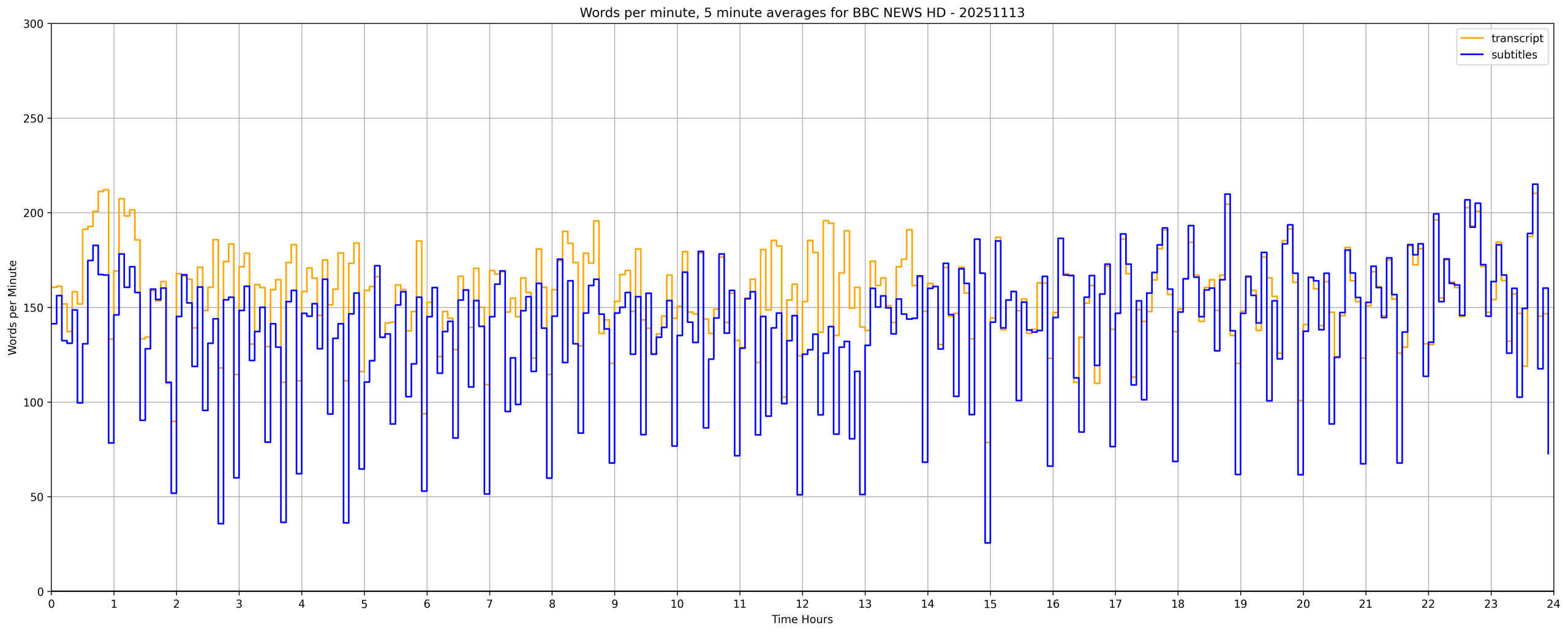
More recently the work has extended to include 24hour plots.

- Delay is plotted as one minute averages.
- Word rate is plotted as five minute averages.
- Programme names are added to the plots.
- And other statistics are logged as a text file.

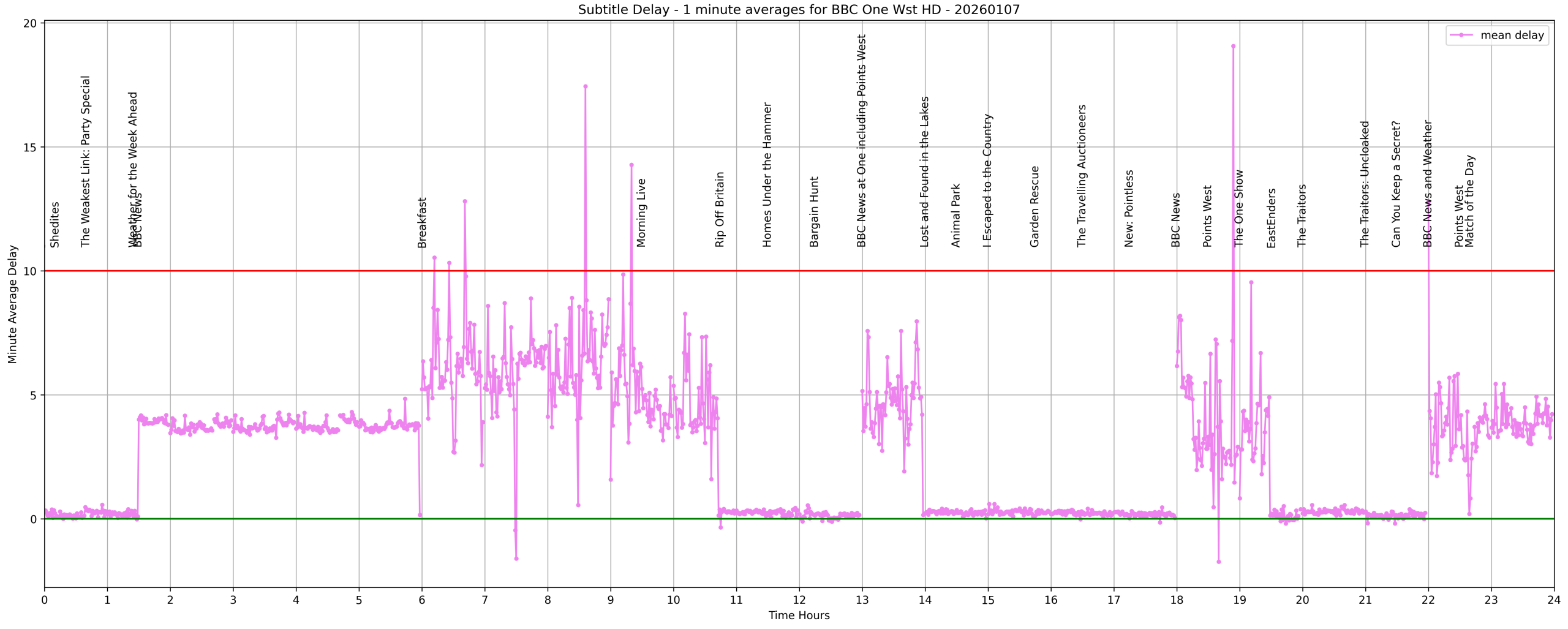
BBC News delay 13th November (Thurs)



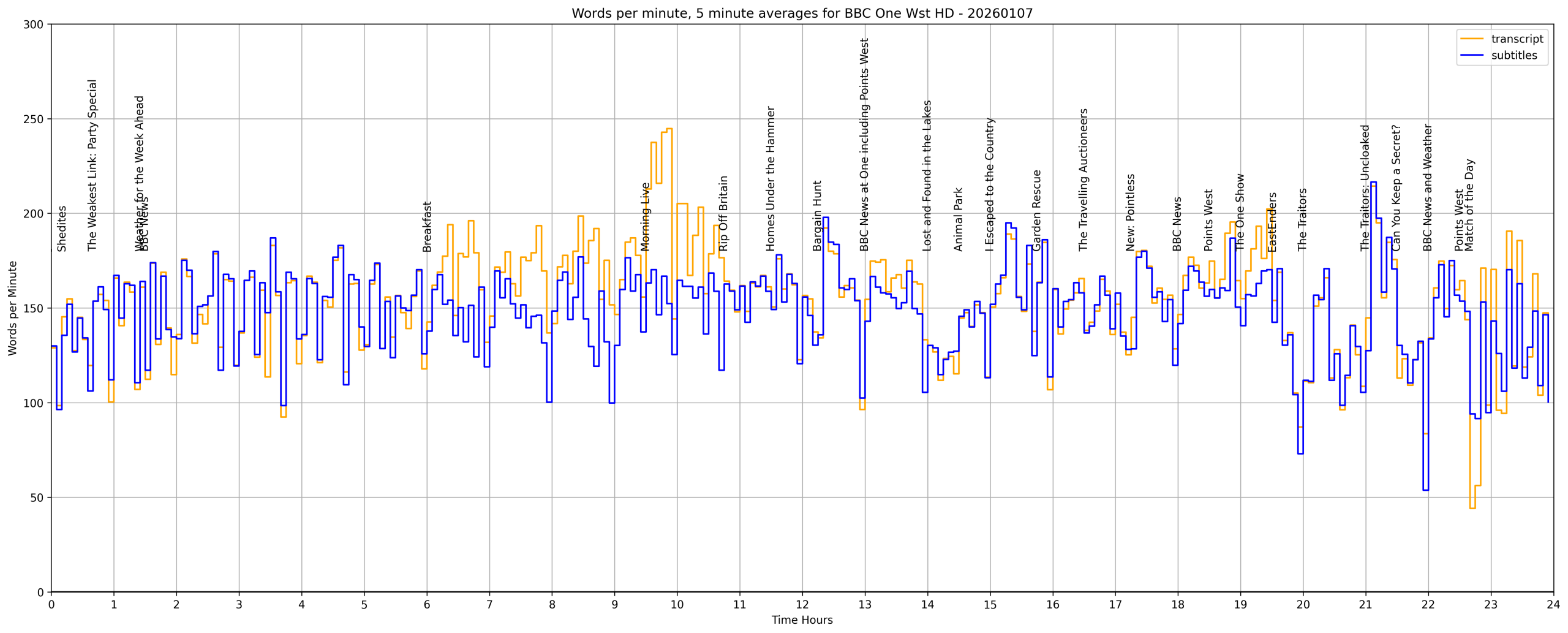
BBC News 5 min word rate 13th Nov



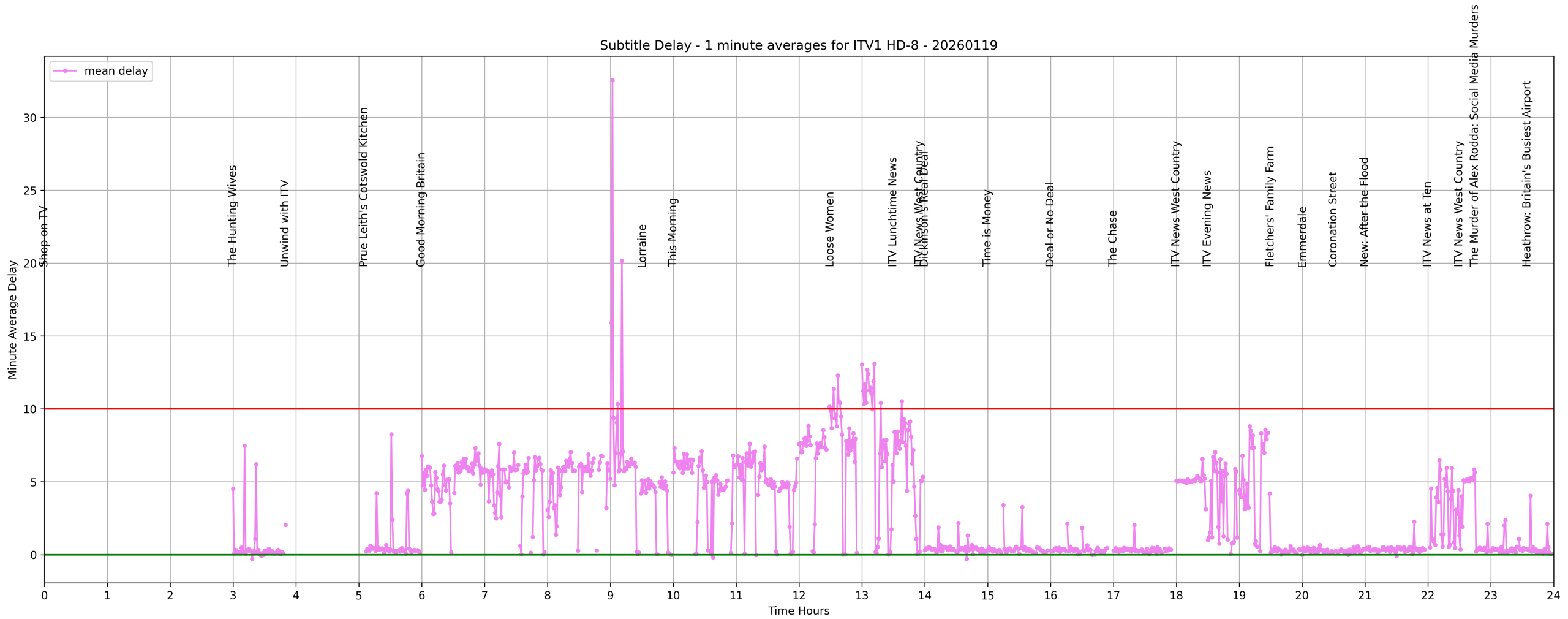
BBC One delay with EPG (Wednesday)



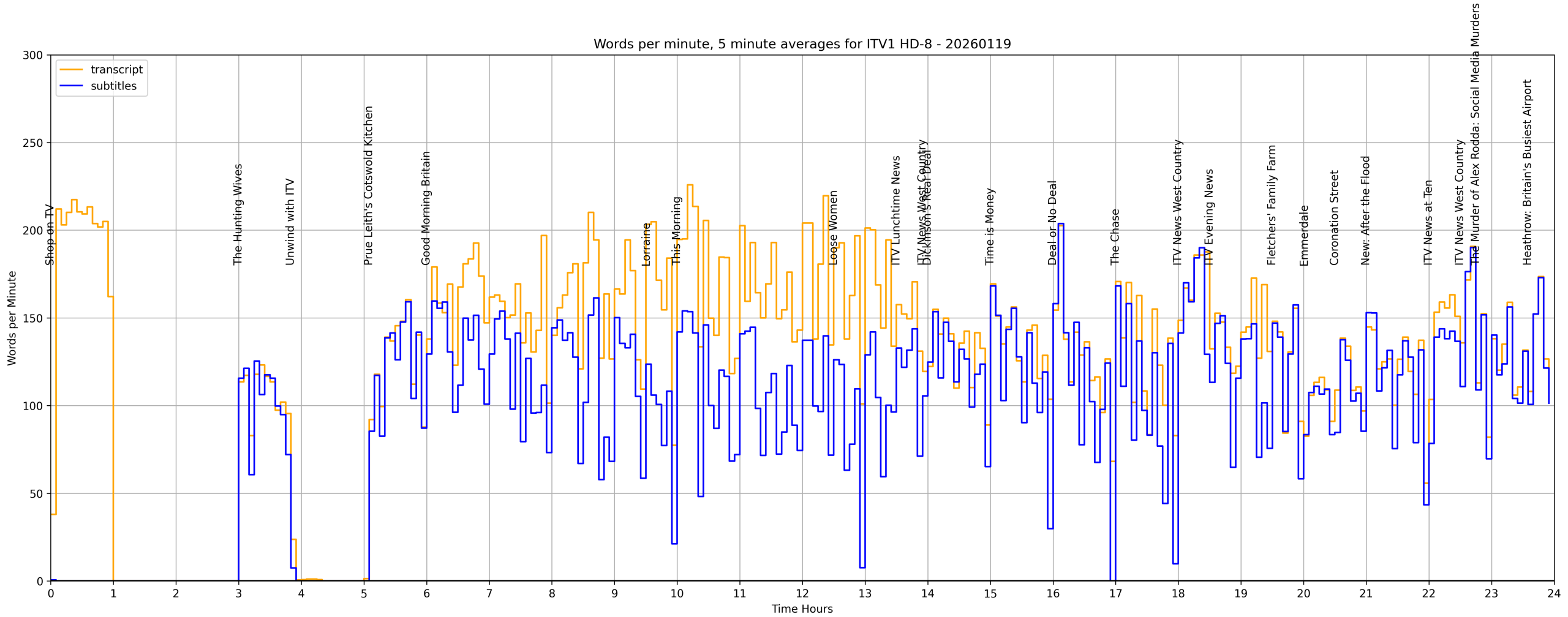
BBC One with EPG - wpm 5 min (Wed)



ITV1 with EPG - delay(Monday)



ITV1 with EPG – wpm 5 min (Monday)



Conclusions

The quality of some TV subtitles is still a significant problem.

This work demonstrates the viability and value of automated subtitle monitoring for delay and word loss.

It is a proof of concept and work in progress.

There are many other issues emerging that have yet to be analysed and documented.

Acknowledgements

I would like to thank

Dr Michael Crabb,

Head of Computing at the University of Dundee and

Dr Carol O'Sullivan,

Associate Professor in Translation Studies at the University of Bristol and co-founder of the UK Subtitling Audiences Network

for their support and encouragement.

For further information visit
www.subtitles.org.uk

IBC2025

USING GENERATIVE-AI SPEECH-TO-TEXT OUTPUT TO PROVIDE AUTOMATED MONITORING OF TELEVISION SUBTITLES

Michael Armstrong

Associate Staff, University of Dundee, UK

ABSTRACT

This paper describes a proof-of-concept approach to monitoring timing errors and word loss in TV subtitles. It reviews previous attempts at subtitle monitoring and the problems caused to viewers by subtitle timing errors and word loss. It then introduces the use of speech to text technology and the conventions in subtitling where repetition, non-speech content and errors can make the task of aligning the speech-to-text transcript to subtitles more challenging. The paper describes the approach taken to remove non-speech content from the subtitles and transcript, along with the natural language processing techniques used to ensure a sufficiently accurate alignment between the two. It then gives examples of the ways in which the results are displayed and some sample results showing the scale of problems with subtitle quality. The paper concludes by reviewing the limits of this approach in terms of accuracy and points out the need for human oversight. Then it goes on to discuss where this approach could be used and other subtitle quality issues which could be monitored automatically.

INTRODUCTION

AI-based speech-to-text tools cannot currently provide broadcast-quality subtitles without human intervention. However, because speech-to-text tools produce different types of errors to those usually found in the production of television subtitling, they can be used to monitor some aspects of subtitle quality. This paper shows how problems with timing and word omission can be detected and quantified by using speech-to-text tools along with natural language processing and simple statistics.

This paper uses the word "subtitles" to refer to text which represents the spoken words along with additional information about the soundtrack in the same language as the speech as provided on UK TV services. These primarily aim to serve D/deaf and hard of hearing audiences but are used by many more people to enhance their viewing experience. This service is also known as "closed captions" outside of the UK. This paper does not cover subtitles for translation which is a more complex topic with different quality issues. The two display modes of subtitling are referred to as "block" where each subtitle appears as sets of words and are replaced by the next subtitle and "snake" where words are added to a subtitle one (or more) at a time and lines scroll up to make space for the next line.