



# Measuring live subtitling quality

Results from the first sampling exercise

Publication date:

30 April 2014



# Contents

Section		Page
1	Summary	1
2	Accuracy	6
3	Latency	9
4	Speed of subtitling	13
5	Edition rates	16
6	Related issues	19
<b>Annex</b>		<b>Page</b>
1	External review of measurements	25
2	Application of the NER model to measuring accuracy	32

## Section 1

# Summary

## Introduction

- 1.1 Ofcom's statutory duties require it, amongst other things, to have regard to the needs of persons with disabilities when carrying out its duties. In particular, Ofcom is required to provide guidance to broadcasters on how they should promote the understanding and enjoyment of their services by people with sensory impairments, including people with hearing impairments<sup>1</sup>.
- 1.2 Given these duties, and continuing complaints about the quality of live subtitling, Ofcom consulted in May 2013 (see footnote 1) on proposals to incentivise broadcasters and access service providers to identify and act upon areas for improvement. Following this consultation, Ofcom published a statement in October 2013 explaining that we would ask broadcasters to measure and report on the key dimensions of the quality of live subtitling, and to report on how many pre-recorded programmes had to be broadcast with live subtitles because they were delivered late.
- 1.3 Shortly thereafter, Ofcom began a two year programme to measure the quality of live subtitling, with a view to identifying areas for improvement and encouraging broadcasters to act upon these. For this purpose, using samples of live subtitling selected by Ofcom from programmes broadcast in October and November 2013, we asked broadcasters to measure the following dimensions of quality<sup>2</sup>:
  - a) the number and type of errors (i.e. minor spelling errors, major omissions or factually misleading subtitles);
  - b) the average latency of the subtitling (the delay between speech and live subtitling), and the range of latencies; and
  - c) the average speed of the subtitling.
- 1.4 These measurements were then checked for consistency by external reviewers from the University of Roehampton.
- 1.5 Sections 2, 3 and 4 of this document report on the results for accuracy, latency and speed, as well as specific points made by the external reviewers. The data was also used to assess the extent to which subtitlers edited down speech, and the results are summarised in section 5. Ofcom considers that the results of this analysis may be of relevance to the debate about the merits of near verbatim vs. edited subtitling. The external reviewers' report also contains some observations on this issue.
- 1.6 In section 6 of this document, we report on related issues, including:

---

<sup>1</sup> Ofcom's relevant powers and duties are described in more detail in section 2 of its consultation document: *The quality of live subtitling – improving the viewer experience*, Ofcom, May 2013 (<http://stakeholders.ofcom.org.uk/binaries/consultations/subtitling/summary/subtitling.pdf>)

<sup>2</sup> *Measuring the quality of live subtitling*, Ofcom, October 2013 (<http://stakeholders.ofcom.org.uk/consultations/subtitling/statement>)

- a) the use of block subtitling in live programming where possible, and the reversioning of subtitling for repeated programmes first broadcast with live subtitling;
- b) the number of pre-recorded programmes that had to be subtitled live (with a consequential reduction in quality) between July and December 2013 because they were delivered too late for pre-prepared subtitling, and the reasons for late delivery;
- c) technical or other failures that occurred between November 2013 and January 2014, resulting in the need for live subtitling of pre-recorded programmes (with a consequential reduction in quality); and
- d) broadcasters' responses to Ofcom's invitation to experiment with inserting short delays in some 'live' programmes which are not time-sensitive, in order to see whether this would enable the quality of subtitles to be improved significantly.

## Key points

### Sampling of live subtitling

- 1.7 The first set of samples, drawn from live-subtitled programmes broadcast in October and November 2013 included extracts from three genres of programming - news, entertainment and chat shows – though not all broadcasters had relevant programming in each genre.
- 1.8 Broadcasters were asked to carry out the measurements themselves, using the NER model<sup>3</sup>, which has been used for some years by academics, broadcasters and subtitle providers to categorise and measure errors. This in turn provided the data that could be used to measure both the speed and latency of subtitling. The measurements were reviewed for consistency and accuracy by a team from the University of Roehampton, and the team's report is at Annex 1.

### Results from the first set of samples

- 1.9 It is important to note that measurements taken on the basis of one set of samples do not provide the basis for drawing firm conclusions. Further samples will be taken over the next 18 months, with the next exercise commencing in May 2014.
- 1.10 In the first set of samples, **accuracy** was generally good, but rather variable. Accuracy rates by genre and broadcaster are shown in the graphs in section 2. Accuracy varied significantly by genre, and was generally higher for news than 98%. However, for chat shows, a number of broadcasters only just achieved overall accuracy rates above 98%, and sometimes dipped well below. The external reviewers suggest that 98% is the quality threshold above which the quality of subtitling can be considered as 'acceptable' (section 2).
- 1.11 For all broadcasters, **latency** in the first set of samples was generally worse than the maximum recommended 3 seconds in Ofcom's guidelines. Typically, it was 75% more (the median latency was 5.6 seconds), with spikes of up to 24 seconds. In most cases, long delays caused omissions and misreporting of information and facts, and

---

<sup>3</sup> The NER model was developed by Pablo Romero-Fresco and Juan Martinez. An explanation of how the model works is given in section 2 of this document.

loss of references to the images (photos or videos) appearing on the screen at the time of the original speech (section 3).

- 1.12 With rare exceptions, the **speed** of the samples measured did not exceed the maximum recommended range of 160-180 words per minute for pre-recorded subtitling (section 4).

## Related issues

### Greater use of block subtitling in live programmes

- 1.13 Following research<sup>4</sup> suggesting that block subtitling is much easier for people to read than scrolling subtitles, broadcasters have made significant efforts to make greater use of block subtitling for pre-recorded inserts in news, and to reversion repeats originally broadcast with live subtitling. Ofcom welcomes this, and will ask for a progress report later in the year.

### Incidence of live subtitling for late-delivered pre-recorded programmes

- 1.14 One way of reducing the problems associated with live subtitling is to avoid it where possible. Consequently, Ofcom asked broadcasters last year to report on how many pre-recorded programmes had to be subtitled live between July and December 2013. The overall results reflected the fact that some broadcasters, such as Sky, Channel 4 and Channel 5 show very few programmes that are produced close to transmission. Others, such as the BBC and ITV, have a higher proportion of topical programming, and face the technical challenge of dovetailing national and regional programming. Thus, while Channel 4, Channel 5 and Sky reported relatively few cases of pre-recorded programming that had to be subtitled live, they were more numerous for ITV (53) and the BBC (155). Many of these were popular peak-time programmes, and so would have affected a significant number of subtitle users. In a number of cases, live subtitling was necessary because production teams were not able to get a copy of the programme to subtitling providers in time.
- 1.15 Ofcom remains concerned that subtitling may not be treated in all cases as an integral part of the production process, even though it is just as important to hearing impaired viewers as, say, matching voiceovers to pictures is to others. The result for viewers who rely upon subtitling is that the quality of their viewing experience is inferior to programmes with pre-prepared subtitles. For this reason, Ofcom will ask broadcasters to submit a further report covering the period July to December 2014. We will publish the results in Spring 2015.

### Technical failures

- 1.16 In order to facilitate a better understanding of how technical failures affect the provision of subtitling (whether live or pre-recorded), Ofcom asked broadcasters to report on how many such failures resulted in the loss of subtitling in the three months from November 2013 to January 2014. Given that technical problems may involve many different parties, and in order to encourage frank reporting, Ofcom said that it would not identify the individual broadcasters concerned.

<sup>4</sup> See, in particular, paragraphs 3.16-3.17 and 3.21–3.24 of *The quality of live subtitling – Improving the viewer experience*, Ofcom, May 2013 (<http://stakeholders.ofcom.org.uk/binaries/consultations/subtitling/summary/subtitling.pdf>)

- 1.17 Broadcasters reported over 400 technical failures leading to temporary loss of subtitling, bearing out viewers' complaints that the problem does not always lie in their receivers. Ofcom notes that there was no obvious pattern to most technical and operational errors. While very infrequent in the context of overall provision, the failures do suggest some relatively common problems which might bear scrutiny by broadcasters, such as subtitles that freeze, subtitling files that were incorrectly coded or flagged, and software glitches.
- 1.18 For this reason, Ofcom will ask for a further report covering the period from November 2014 to January 2015. The anonymised results will be summarised as part of the report on live subtitling due in Spring 2015.

### Short delays to improve subtitling quality

- 1.19 In its May 2013 consultation (see footnote 2) Ofcom invited consultees to tell us about the factors that might facilitate or hinder the insertion of a delay in live transmissions sufficient to improve the quality of subtitling. Broadcasters were strongly opposed to inserting delays, citing the risk to viewer trust, the need for complex technical solutions, and competition from other media, amongst other reasons.
- 1.20 In its statement of October 2013 (see footnote 1), Ofcom noted the potential improvements to the quality of live subtitling to be gained from short delays in live transmissions (e.g. synchronisation and block subtitles), but noted too the strong views of broadcasters that the detrimental impact on viewer trust, as well as the operational risks, made this 'wholly disproportionate'.
- 1.21 Ofcom said that it would discuss further with broadcasters whether, in some programmes that were not time-sensitive, delays could be justified on editorial grounds to protect viewers from the certainty of poorer-quality subtitles.
- 1.22 In response, broadcasters reiterated their opposition to experimenting with short delays in transmission to improve the quality of subtitling, citing the same reasons as previously. However, they continued to use such delays for other editorial reasons (e.g. the Pistorius trial). As the reasons for distinguishing between the interests of hearing-impaired viewers and viewers in general are not clear, Ofcom will seek further discussions with broadcasters.

### **Next steps**

- 1.23 The next sampling exercise will begin in May 2014, drawing on news, chat show and entertainment programmes broadcast in April and May 2014 with live subtitles. We plan to report on the outcome in Autumn 2014. This report will include an update on progress by broadcasters in the use of block subtitling in live programmes.
- 1.24 The following round, scheduled for October 2014, will sample programmes broadcast with live subtitles in October and November 2014. We plan to report on the outcome in Spring 2015. This report will include updates on the number of pre-recorded programmes that were delivered late and so had to be subtitled live during the period July to November 2014, and the number of technical failures resulting in the loss of subtitling during the period November 2014 to January 2015.
- 1.25 The final set of samples will be taken in May 2015. We expect to report on the results in Autumn 2015. This final report will pave the way for consideration of any future action that may be indicated, including a review of Ofcom's guidance.

- 1.26 Ofcom would like to thank both broadcasters and subtitling providers for their collaboration in this project, as well as the external reviewers from the University of Roehampton.

## Section 2

# Accuracy

## Why accuracy is important

- 2.1 Ofcom's May 2013 consultation noted that hearing-impaired viewers were concerned about the accuracy of subtitling. While some mistakes are inevitable, for example when homophones of the right word are mistakenly shown, such as 'their' in place of 'there' or 'they're', they force viewers to spend time working out what was meant. In more serious cases, subtitles may be so garbled as to be unintelligible. Worse still, in some situations, a plausible error (such as the substitution of '15%' for '50%') may lead the viewer to believe that they understand the meaning, when they are actually being misinformed.

## How accuracy is measured

- 2.2 Broadcasters were asked to carry out the measurements themselves, using the NER model, which has been used for some years by academic, broadcasters and subtitle providers to categorise and measure errors. This in turn provided the data that could be used to measure both the speed and latency of subtitling. The measurements were reviewed for consistency and accuracy by a team from the University of Roehampton, and the team's report is at Annex 1.
- 2.3 Accuracy rates are calculated according to the NER Model<sup>5</sup>, which was developed by a team of researchers at the University of Roehampton. The same team identified 98% as the quality threshold above which the quality of subtitling can be considered as 'acceptable'. The NER model uses the following formula:

$$NERvalue = \frac{N - E - R}{N} * 100$$

- 2.4  $N$  represents the total number of words in the sample of live subtitling.  $E$  represents the value of edition errors, and  $R$  represents the value of recognition errors. Minor edition or recognition errors (where the meaning is easily understandable) are scored 0.25, 'normal' or 'standard' errors (where it is obvious that there is an error, but the intended meaning is difficult or impossible to understand) are scored 0.5, and serious errors (where there is an error, but this may not be apparent to the subtitle reader, e.g. the use of 15% instead of 50%) attract a score of 1.
- 2.5 Annex 2 illustrates how the model has been applied in practice, using excerpts from three news programmes.

## Summary of results

- 2.6 Figures 1, 2 and 3 below show overall accuracy rates in the samples taken from news bulletins, chat shows and entertainment programmes respectively, for all broadcasters, and for individual broadcasters.

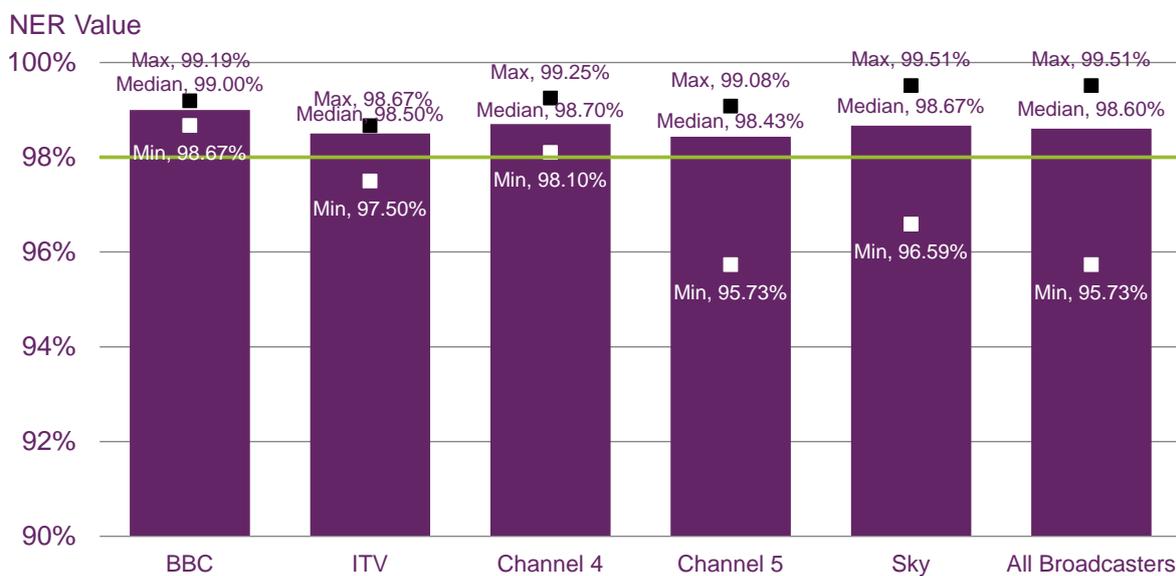
---

<sup>5</sup> Pablo Romero- Fresco and Juan Martínez, *Accuracy Rate in Live subtitling – the NER Model* (2011), Roehampton University, UK (<http://roehampton.openrepository.com/roehampton/bitstream/10142/141892/1/NER-English.pdf>)

2.7 The external reviewer noted that, the further accuracy rates fell, the more problems they presented. They suggested that the present analysis confirms the figure of 98% as a valid threshold from which subtitles may be considered of acceptable quality<sup>6</sup>. For reference purposes, the graphs below show the 98% threshold with a bold green line.

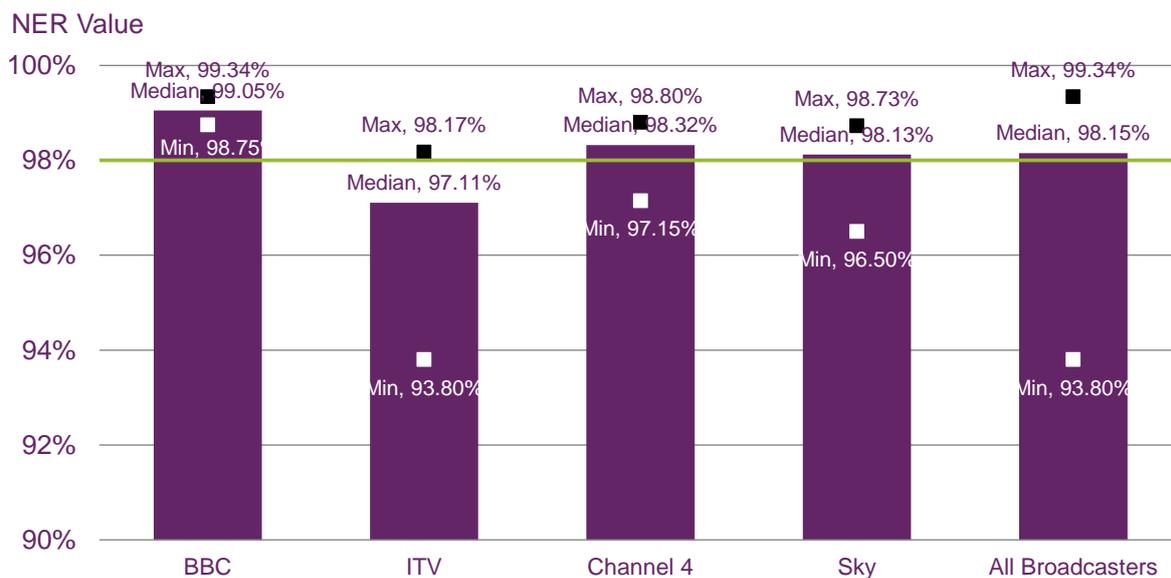
2.8 In this sample, and taking into account other factors related to overall quality, subtitles with 99%-99.5% accuracy rate ranged from very good to excellent, whereas those with 98.5%-99% ranged from good to very good. In two of the programmes analysed, accuracy rates below 98% were found to be acceptable on the basis of the other elements in the subtitles.

**Figure 1: accuracy rates in samples drawn from news bulletins**  
(higher bars represent a better viewing experience)

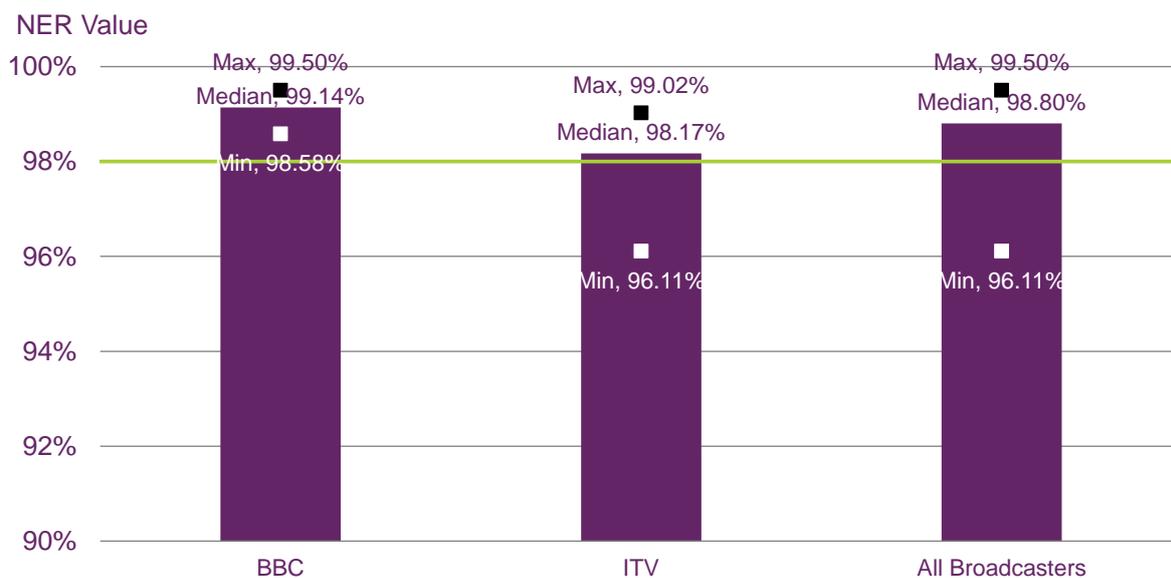


<sup>6</sup> 98% may seem a high threshold, but could be achieved by subtitles with some serious errors, or many minor errors, or a combination of both, as the examples in Annex 2 show. This threshold has been suggested as appropriate by Pablo Romero- Fresco and Juan Martínez, *Accuracy Rate in Live subtitling – the NER Model* (2011), Roehampton University, UK.

**Figure 2: accuracy rates in samples drawn from chat shows**  
 (higher bars represent a better viewing experience)



**Figure 3: accuracy rates in samples drawn from entertainment programmes**  
 (higher bars represent a better viewing experience)



## Section 3

# Latency

## Why latency is important

- 3.1 The term 'latency' is used in this context to denote the delay between speech and subtitling. Ofcom's May 2013 consultation said that many viewers found latency one of the most frustrating aspects of live subtitling. Latency forces viewers to relate subtitles to images that may have disappeared from the screen, resulting in a disjointed viewing experience. The longer the delay, the less satisfactory the viewing experience.
- 3.2 Ofcom's guidelines recommend that the delay in subtitling presentation on live programmes should be no greater than 3 seconds. Ofcom has said that it will review this guidance in the light of the data gathered from the two year quality measurement exercise.

## How latency is measured

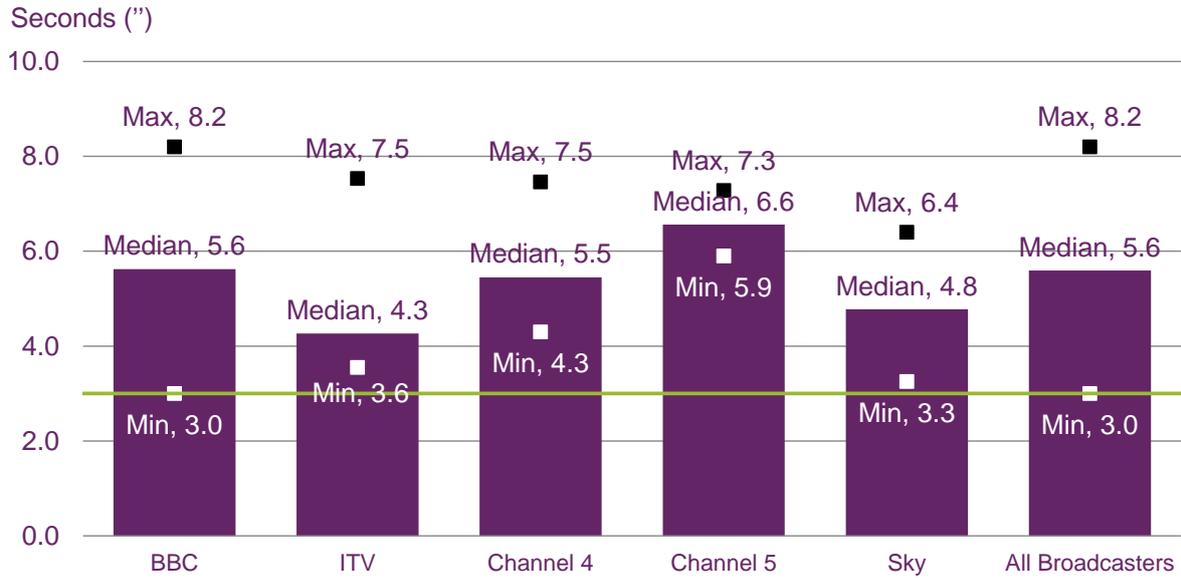
- 3.3 To measure latency, broadcasters were asked to identify 2-3 words per minute in the audio at intervals of 20-30 seconds and to measure the delay between the spoken word and the subtitles. As the latency of scrolling subtitles may vary depending on what word is measured, broadcasters were asked to ensure that 50% of the words identified appeared first in a subtitle, and the other 50% last in a subtitle.
- 3.4 Typically, although not always, the last words of a long subtitle may be more delayed than the first ones if there is a change of speaker, if there is overlapping or very fast speech, or if there are errors in the subtitles that have been corrected live by the subtitler. In order to measure the average latency of scrolling subtitles is thus important to take into account both the beginning and the end of the subtitles.
- 3.5 So, each one of the programmes in the sample had numerous measurements of latency, and these were averaged to obtain a unique measure per programme. But it should be noted that, although for reasons of practicality this analysis uses the averages of the delays in the subtitling, the range of latency in most programmes was quite significant.

## Summary of results

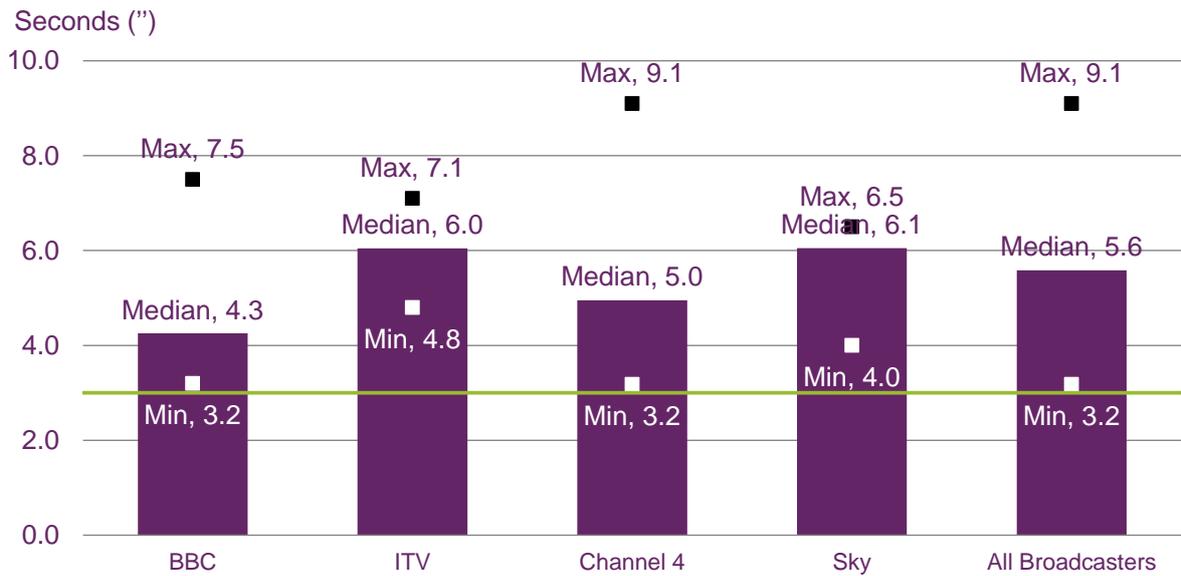
- 3.6 Figures 4, 5 and 6 below show overall latency in the samples taken from news bulletins, chat shows and entertainment programmes respectively, for all broadcasters, and for individual broadcasters. For reference purposes, the recommended maximum 3 second delay is marked on the graphs below with a bold green line.
- 3.7 In specific comments, the external reviewers noted that:
- a) the median latency of the programmes analysed was 5.6 seconds. As shown in the sample analysed here, the delay of live subtitles depends largely on the availability and use of pre-prepared scripts, the live subtitling technique used, the genre of the programmes and the occurrence of specific technical issues during transmission;

- b) when scripts are available and pre-prepared subtitles are combined with live subtitles, the average delay has been reduced to 3-4 seconds and on rare occasions even to 2-3 seconds. Subtitles made by stenographers have less delay (an average of 3-5 seconds) than those made by respeakers, which are closer to 6 and often 7 seconds;
- c) on the whole, programmes with many speakers, fast speech rates and overlapping speech are likely to cause delays of over 7 seconds, which can have a significant impact on viewers' comprehension;
- d) finally, regardless of the average delay of the subtitles and possibly due to specific technical problems, several programmes in the sample analysed featured peaks of 10-second delays and sometimes even more. In some cases, these peaks reached delays of 15, 17, 18 and even 24 seconds. In the majority of cases, these long delays caused omission and misreporting of information and facts, rewording of sentences in a way that became misleading and loss of references to the images (photos or videos) appearing on the screen at the time of the original speech.

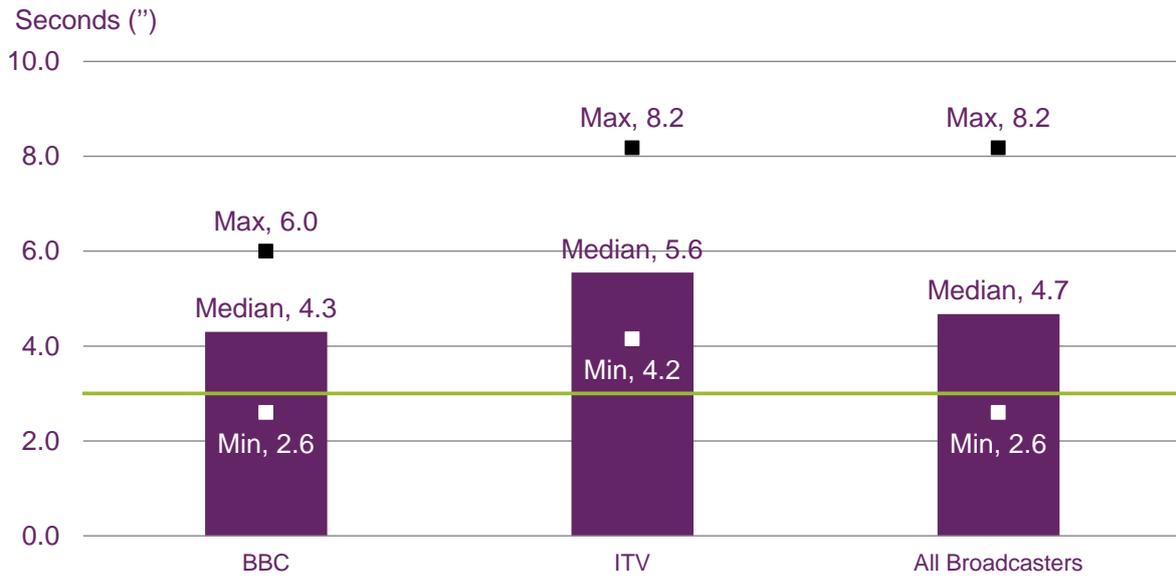
**Figure 4: latency in samples drawn from news bulletins**  
 (higher bars represent an inferior viewing experience)



**Figure 5: latency in samples drawn from chat shows**  
 (higher bars represent an inferior viewing experience)



**Figure 6: latency in samples drawn from entertainment programmes**  
(higher bars represent an inferior viewing experience)



## Section 4

# Speed of subtitling

## Why the speed of subtitling is important

- 4.1 For some years, it has been widely acknowledged that the speed of subtitling can affect comprehensibility. For this reason, Ofcom's guidelines recommend that 'the speed should not normally exceed 160 to 180 words per minutes for pre-recorded programmes. Although it may not be practicable to restrict the speed of subtitles for all live programmes, commissioning editors and producers should be aware that dialogue which would require subtitles faster than 200 wpm would be difficult for many viewers to follow'<sup>7</sup>.
- 4.2 In its May 2013 consultation document, Ofcom pointed to research suggesting that the speed of subtitling (as measured in 'words per minute' or "wpm") has an impact both on the levels of comprehension, and on the amount of time viewers are able to watch what is happening on screen, as opposed to reading subtitles.
- 4.3 In its October 2013 statement, Ofcom concluded that there would be merit in looking again at its guidance on the recommended maximum speed of subtitling in the light of the research now available, but said that it would be better to do so having regard to data from the two year measurement project. In doing so, Ofcom will have also regard to the views of interested parties.

## How the speed of subtitling is measured

- 4.4 The speed of subtitling is usually measured in words-per-minute (wpm) rates. In order to calculate the wpm rate, broadcasters were asked to sum the words in each segment and divide this total by the duration of the segment.

## Summary of the results

- 4.5 Figures 7, 8 and 9 below show the overall speed of subtitling in the samples taken from news bulletins, chat shows and entertainment programmes respectively, for all broadcasters, and for individual broadcasters. For reference purposes, the upper and lower boundaries of the maximum range recommended for pre-recorded subtitling is indicated on the graphs below with bold coloured lines.
- 4.6 The speed of most samples of live subtitling was below the recommended maximum band of 160-180 wpm. Only in three instances was the subtitling speed higher than 180 wpm.
- 4.7 In specific comments, the external reviewers noted that:
- a) entertainment programmes had the lowest average speech rate (161wpm). Such programmes combine live spontaneous dialogue with pre-recorded material and songs (thus providing much-needed gaps for subtitlers) and they lend themselves

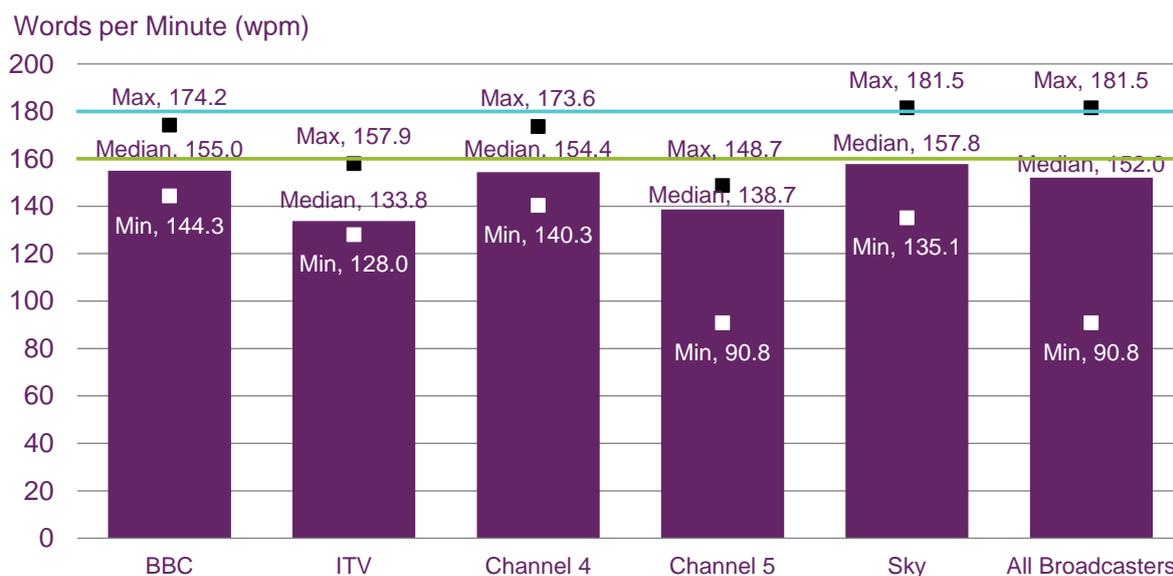
---

<sup>7</sup> Ofcom, *Guidelines on the provision of access services*, Annex 4 to Code on Television Access Services (<http://stakeholders.ofcom.org.uk/binaries/broadcast/other-codes/tv-access-services-2013.pdf>).

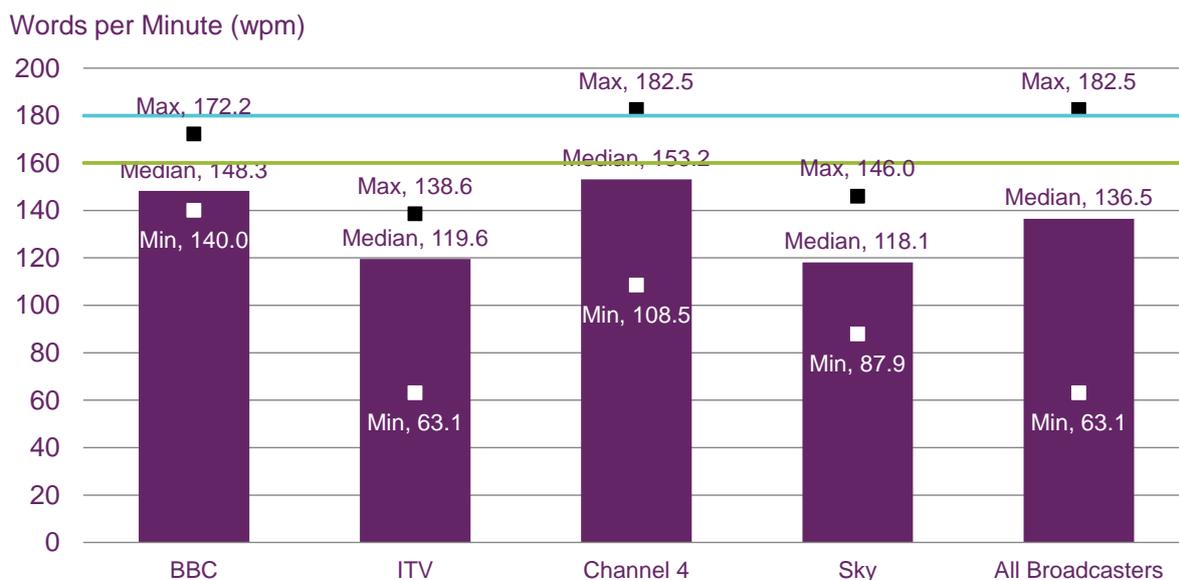
to a certain degree of editing, since the spontaneous dialogue often features hesitations, false starts and redundant comments;

- b) news programmes exhibited a similar average speech rate (163wpm) to entertainment programmes but they normally have more serious content, which is particularly challenging to edit without losing important information; and
- c) chat shows are often the most difficult genre as far as live subtitling is concerned, not only because of the speakers' fast speech rates (an average of 180wpm in this sample, with peaks of 250wpm) but also because of overlapping speech that is essential to understand the humour in the programme.

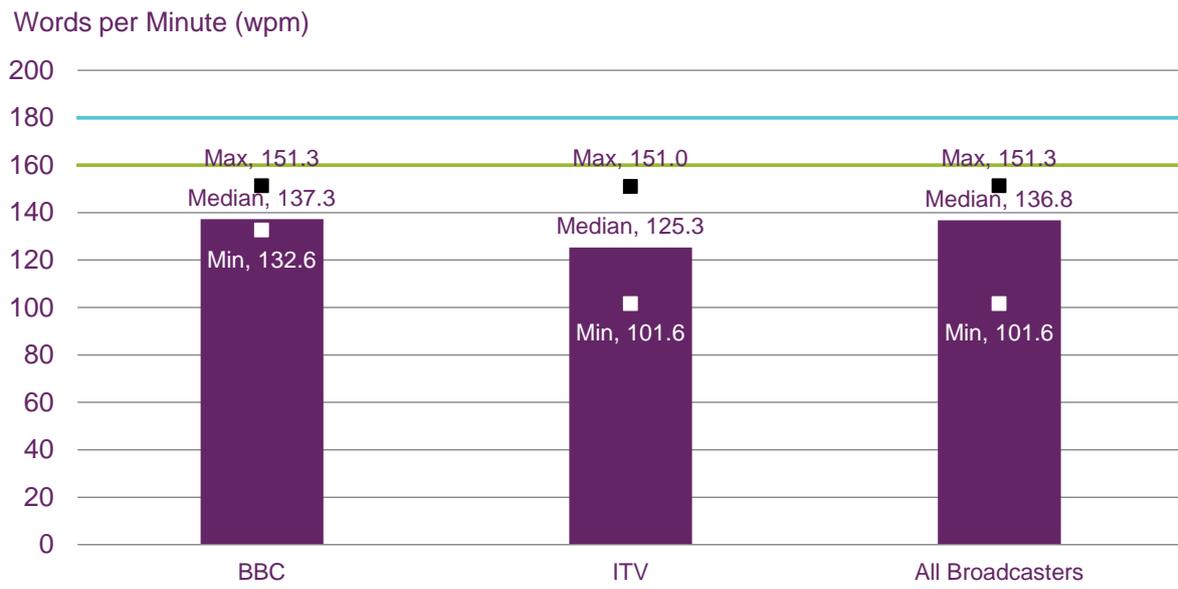
**Figure 7: speed of subtitling in samples drawn from news bulletins**



**Figure 8: speed of subtitling in samples drawn from chat shows**



**Figure 9: speed of subtitling in samples drawn from entertainment programmes**



## Section 5

# Edition rates

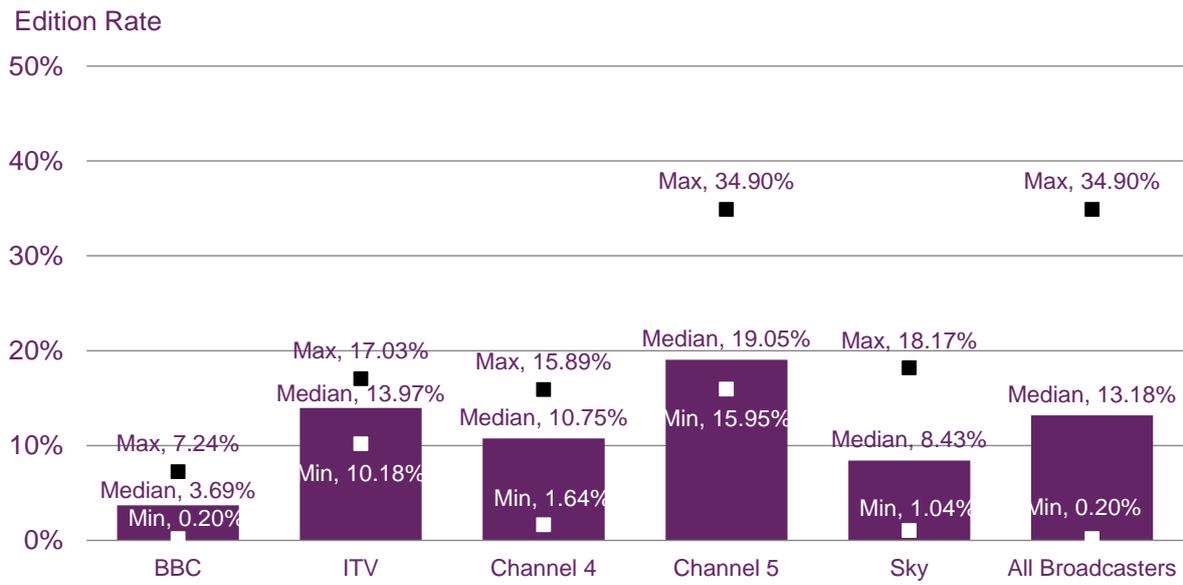
## Why editing is needed

- 5.1 The speed at which people speak on television (i.e. 'speech rate') can vary significantly, and is sometimes faster than the rate at which subtitles can be created by respeakers or even easily read by users. For these reasons, subtitle providers often have to edit what is said for the purpose of subtitling.
- 5.2 The data collected for this exercise allowed us to calculate the extent to which the subtitles represented an edited version of what was said – the so-called 'edition rate'.
- 5.3 The edition rate is not a measure of quality, and there is no guidance on appropriate edition rates. Nevertheless, we consider that the information on how much speech is edited out will help to inform the debate between those seeking near verbatim subtitling, and those who advocate greater amounts of editing in order to keep the speed of subtitling down.

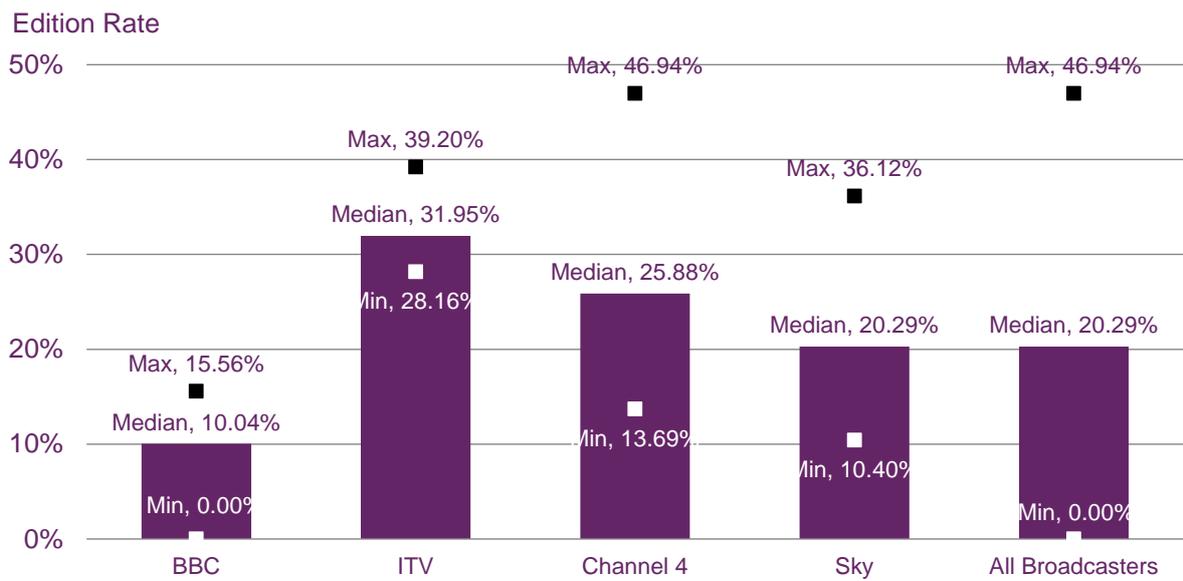
## Summary of results

- 5.4 Figures 10, 11 and 12 summarise the overall edition rates by genre and broadcaster. The average edition rate in the whole sample was 16.42%, and the median was 15.14%.
- 5.5 The external reviewers also looked at edition rates, and their comments are set out in Annex 1. It is important to note that Annex 1 sets out average edition rates, and so differ slightly from the median rates reported in this section. In specific comments on edition rates in different genres, the external reviewers (Annex 1) noted that:
  - a) 'given the importance of [news] programmes and their content, subtitlers seem to be making a significant and commendable effort to edit less than in entertainment programmes (an average editing rate of 13%), managing to keep the accuracy rate above the required threshold (98.60%)';
  - b) as regards chat shows, 'subtitlers have thus been forced to edit an average of 30% of the original content and, on the whole, have just managed to reach the accuracy threshold (98.15%)'. This reflected 'speakers' fast speech rates (an average of 180wpm in this sample, with peaks of 250wpm) but also ... the presence of overlapping speech that is essential to understand the humour in the programme'; and
  - c) entertainment programmes 'lend themselves to a certain degree of editing, since the spontaneous dialogue often features hesitations, false starts and redundant comments. This may explain why despite a fairly high editing rate (an average of 20%), the entertainment programmes in the sample have obtained the best overall accuracy rates of all three genres (98.80%)'.

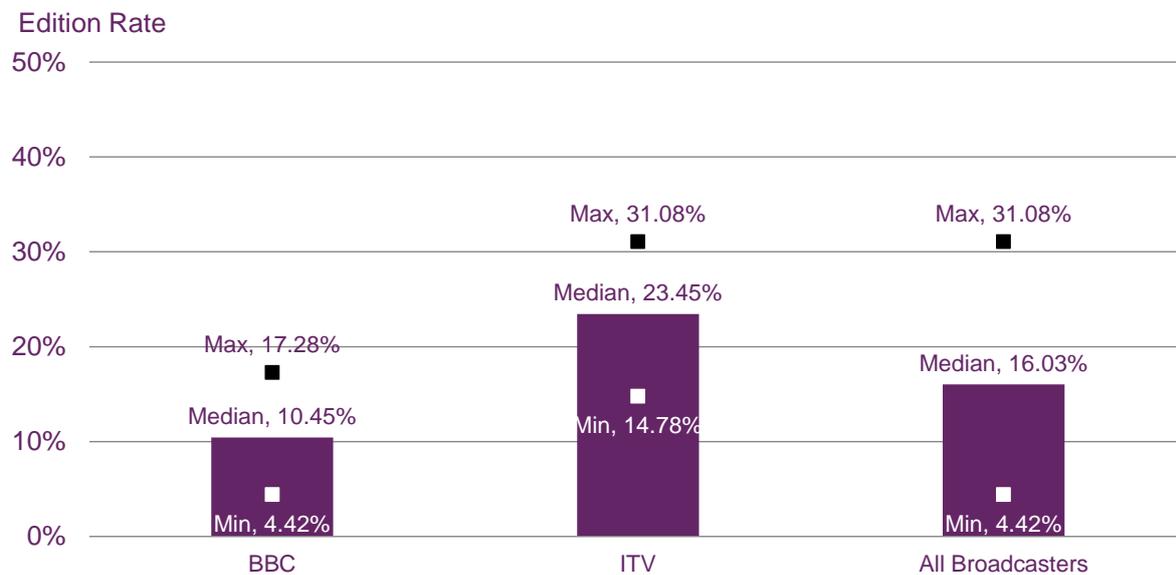
**Figure 10: edition rate in samples drawn from news bulletins**



**Figure 11: edition rate in samples drawn from chat shows**



**Figure 12: edition rate in samples drawn from entertainment programmes**



## Section 6

# Related issues

## Greater use of block subtitles in live programmes

- 6.1 In the light of evidence reported in Ofcom's May 2013 consultation that block subtitling was much easier for viewers to read and comprehend, Ofcom's statement of October 2013 encouraged broadcasters to use block subtitling wherever possible, and strongly recommended that repeats of programmes originally broadcast with scrolling subtitles be reversioned with block subtitles.
- 6.2 In response, all broadcasters indicated a willingness in principle to use block subtitles wherever possible:
- a) the BBC said that it was now able to transmit a mix of block and scrolling subtitles, and was testing the use of block subtitles, initially in nations and regions news programmes, where pre-recorded inserts were available and when material was repeated. The aim was to implement these arrangements more widely in the coming months;
  - b) ITV said that it was trialling block subtitles in place of scrolling subtitles in regional news bulletins in the Granada region, and on the ITV network programme *This Morning* on 19 February. Feedback was being provided with help from external reviewers identified by Action on Hearing Loss, and ITV hoped to report soon on the results. ITV also said that its subtitling provider had been reversioning repeats with block subtitles for several years;
  - c) Channel 4 said that it always looks to use block subtitles whenever possible, both in the original programmes and in repeats of programmes in place of the original scrolling subtitles;
  - d) Channel 5 said that it had substantially increased the use of block subtitles, as was evident from its daily news programmes. It had also introduced new procedures to enable subtitles to be edited and tidied up on next day repeats of *Big Brother* and *Celebrity Big Brother*, which included much greater use of block subtitling. Apart from these programmes, Channel 5 had very few shows that were subtitled live and then repeated; and
  - e) Sky said that its aim is to use blocks wherever possible on live programmes, so it is now using live blocks on all Sky News and Sky Sports News programmes where it is able to prepare material in advance – this can either be from newsroom scripts or VTs which have been aired previously. It pointed out that exceptions could occur with breaking news or where the subtitler was not yet using the latest software. However, the vast majority of the time there would be live blocks on those channels. Sky said that it was also using live blocks on repeated programmes wherever possible, and provided a list of recent examples.
- 6.3 Some of the samples of news programmes measured in the recent exercise included extracts combining live and pre-prepared subtitles. The external reviewer noted that the availability of scripts for some portions enabled synchronised verbatim subtitles to be produced. However, there were difficulties at the junction between live and pre-prepared subtitling, which often involves a considerable delay in the live part or an excessive speed when live subtitles give way to pre-prepared subtitles so that they

can catch up with speech. One suggestion was that pre-prepared blocks could be introduced gradually with some delay until they are completely synchronous.

- 6.4 Ofcom will invite broadcasters to provide updates in time for these to be included in the second report on subtitling quality, due in Autumn 2014.

## **Incidence of live subtitling for late-delivered pre-recorded programmes**

- 6.5 Ofcom's consultation drew attention to the problems that occurred when pre-recorded programmes were delivered to broadcasters too late to allow subtitling to be prepared in advance, necessitating lower quality live subtitling. We asked broadcasters to let us know which late-delivered programmes had to be subtitled live between July and December 2013.
- 6.6 The main reasons for live subtitling of pre-recorded programmes that were delivered too late for pre-prepared subtitling were as follows:
- a) late edits: some programmes had to be edited close to transmission for legal and editorial reasons, which delayed release of the final version to subtitling providers. Some of these were topical programmes that have to be produced close to transmission;
  - b) late deliveries: some in-house and external producers were unable to complete the production and have it reviewed for technical compliance in time to permit pre-prepared subtitling; and
  - c) technical problems: on occasion, technical problems, such as corrupted files, caused insuperable delays.
- 6.7 In other cases, some pre-recorded programmes had to be live-subtitled because difficulties encountered in the delivery of programmes from regional offices to the subtitling service provider, or because the repeat of a live programme was too close to the original for the timely creation of a pre-recorded subtitle file.
- 6.8 The overall results reflected the fact that some broadcasters, such as Sky, Channel 4 and Channel 5 show very few programmes that are produced close to transmission. Others, such as the BBC, have a higher proportion of topical programming as well as facing the technical challenge of dovetailing national and regional programming:
- a) the BBC reported that 155 late delivered programmes had to be subtitled live, of which 51 were shown in peak time. A significant proportion of these (73 programmes, or 47%) were pre-recorded topical programmes that were delivered late, such as *Mock the Week*, *Have I Got News For You*, the *Graham Norton Show* and *Top Gear*. Others, such as *Later ... with Jools Holland*, *Sweat the Small Stuff*, *The Blame Game* and *Friday Download* (CBBC) were delivered late by the production team (66, or 43%). A further six programmes produced in the nations were delivered late from BBC nations' offices. The BBC pointed out that late-delivered programmes accounted for just 0.51% of broadcast hours;
  - b) ITV broadcast 53 late-delivered programmes with live subtitling. Several involved *The Jeremy Kyle Show*, where delays resulted from late edits and other production problems which ITV has asked the producer to address. In addition, there were some late substitutions of programmes due to the illness of the presenter. Of other delays, a few were sports highlights programmes produced

on the same day, and some resulted from the need to check the content for compliance with the Ofcom Broadcasting Code rules. A few were programmes delivered late by the production team;

- c) Channel 4 broadcast six late-delivered programmes with live subtitles, most of which, such as *999: What's Your Emergency* were edited very late for legal and other reasons. In another case, transmission was brought forward, preventing timely delivery to subtitling providers;
- d) Channel 5 reported that nine pre-recorded programmes had to be subtitled live because of late delivery, of which three were in peak-time. It said that some of the incidents involved only minor inconvenience, such as an increase in latency by one second. In three cases involving *Celebrity Super Spa*, decisions to advance transmission meant that there was insufficient time to prepare subtitling in advance. Technical problems prevented the provision of an advance copy of the programme to subtitlers in some other cases.
- e) Sky reported that none of its programmes in the genres measured were delivered past a contractual deadline and thus had to be subtitled live. For programmes such as *Soccer AM* or *Game of Thrones* that were normally delivered late, its practice was to split them into segments so that multiple subtitlers can work on them, enabling pre-recorded subtitling file to be produced, or subtitles to be cued live in blocks. However, it pointed out that pre-recorded sports events programming (not one of the genres considered in the current exercise) was often delivered to tight timescales, and so live subtitling in these circumstances was considered to be 'business as usual'.

- 6.9 Ofcom understands that broadcasters have worked hard in an attempt to reduce the number of late-delivered programmes that have to be subtitled live, and welcomes this. In some cases, delays are unavoidable. In any case, late-delivered programmes account for a small minority of total programming. Nonetheless, the survey showed that a significant number of pre-recorded programmes (often scheduled for peak-time viewing) still had to be subtitled live, often because the production teams were not able to get a copy of the programme to subtitling providers in time.
- 6.10 Ofcom remains concerned that subtitling may not be treated in all cases as an integral part of the production process, even though it is just as important to hearing impaired viewers as, say, matching voiceovers to pictures is to others. The result for viewers who rely upon subtitling is that the quality of their viewing experience is inferior to programmes with pre-prepared subtitles. For this reason, Ofcom will ask broadcasters to submit a further report covering the period July to December 2014. We will publish the results of this in Spring 2015.

### Short delays to improve subtitling quality

- 6.11 In its statement of October 2013, Ofcom noted the potential improvements to the quality of live subtitling to be gained from short delays in live transmissions (e.g. synchronisation and block subtitles), but noted too the strong views of broadcasters that the detrimental impact on viewer trust, as well as the operational risks, made this 'wholly disproportionate'.
- 6.12 Ofcom recognised the sensitivity around news and results-based programmes, and the practical difficulties around the coordination of different streams of regional programming. Noting that some broadcasters consider short delays editorially justified to safeguard viewers from the possibility that they might hear obscenities,

Ofcom said that it would discuss further with broadcasters whether, in some programmes that were not time-sensitive, delays could be justified on editorial grounds to protect viewers from the certainty of poorer-quality subtitles.

- 6.13 In general, broadcasters remain opposed to inserting delays in live transmissions to improve the quality of live subtitling, even on an experimental basis:
- a) the BBC noted that it had already given careful thought to this suggestion as part of the consultation process, and it remained unconvinced that it would deliver a net benefit to licence fee payers. The length of delay to deliver a proportionately small change would have to be well over 20 seconds, and it would add significant technical and operational risk to the BBC's already complex playout operation, particularly in relation to the junctions between UK and nations/regions news. The BBC said that live TV created shared experiences for its audience, including the deaf community. In order not to mislead the audience, the delayed output would have to be labelled as such, with reasons;
  - b) ITV said that delaying live transmissions to allow better quality subtitles would be a wholly disproportionate measure most unlikely to be in the interests of all viewers. Aside from the significant direct and indirect costs for ITV in adopting such an approach, delays in live transmission would bring a degree of complexity and risk to ITV regions and sub-regions which ITV does not feel could be justified. ITV also believes such an approach would jeopardise viewer trust in ITV programmes and negatively impact audiences' perceptions of ITV;
  - c) Channel 4 remained of the view that the costs and risks associated with doing so would be disproportionate to the potential improvements that could be made to subtitling quality. Most live subtitled content - for example sports content or news programming – was highly 'time-specific' and it would not be appropriate for Channel 4 to apply any transmission delays to them. Channel 4 had considered the feasibility of applying delays to non-live programmes that receive live subtitling – for example topical programmes that are filmed close to transmission. However, it strongly believed that the costs of introducing delays of this kind to these programmes would be disproportionate to any potential benefits, and that it would be unlikely to be in the interests of either hearing impaired viewers, or the wider British viewing public; and
  - d) Channel 5 said that it has considered at length the possibility of introducing additional delays into recorded programmes that have to be subtitled live. However, it had not found any reason to depart from its original view, that introducing a delay would create a new element in the already complex technical process of delivering television pictures to viewers' screens, and would increase the possibility of something going wrong and harming viewers' experience.
  - e) Sky said that it would need to investigate this proposition further to properly consider the technical and operational implications depending on the particular type of output.
- 6.14 Ofcom accepts that short delays in news bulletins would be both controversial and, in the case of regional bulletins, more risky, which is why we suggested that broadcasters experiment with delays in programmes that are not time-sensitive. Ofcom notes that, notwithstanding cost, technical complexity and operational risk, broadcasters such as the BBC and Sky have continued to make use of short delays in the transmission of live events (notably in the recent trial of Oscar Pistorius, where

delays ranged from 30 seconds to one minute). Both Channel 4 and Channel 5 have introduced similar delays in the past, for editorial reasons.

- 6.15 This suggests that there are circumstances in which broadcasters find that the technical and operational obstacles to inserting short delays in live programmes are not insuperable, and that the costs are justifiable, and serve the interests of viewers. Ofcom will discuss with broadcasters why they consider that it would not be in the interests of viewers to insert similar delays in live programmes that are not time sensitive in order to improve the quality of subtitling.

## Technical and other issues

- 6.16 Ofcom referred in the consultation paper to technical and other issues that lead to the loss of or interruption to subtitling, for a variety of different reasons. We said that, in order to understand whether there are particular aspects of the production or transmission processes that require attention, we would be asking broadcasters to provide information on the incidence, severity and causes of failures in the provision of subtitling.
- 6.17 Ofcom asked broadcasters to report on subtitling failures occurring in the three months beginning on 1 November 2013 due to technical and other issues that lead to the loss of or interruption to subtitling. We recognised that, as the processes often involve multiple parties, there might be issues of commercial confidentiality. In order to encourage broadcasters to be as frank as possible in providing information, we said that we would publish the information in confidential form – hence this report does not identify individual broadcasters.
- 6.18 The reasons for subtitling failures fell into three main categories:
- a) operational – all instances in which there have been administrative or operational errors resulting in the partial or complete loss of subtitling to from a programme;
  - b) technical – all situations in which any sort of technical difficulty (such as a problem with the subtitling software) has stopped the normal provision of subtitling; and
  - c) other - all cases which do not relate to any of the other two categories above, such as failures relating to distribution or compliance problems.
- 6.19 Broadcasters reported that, during the three months from November 2013 to January 2014, there were subtitling failures on 437 occasions. Clearly, such failures only disrupted viewing to a very small proportion of all the programmes shown over the period, though such instances are bound to be annoying for those affected.
- 6.20 Key points to note are as follows:
- a) in most cases where the duration of outages was reported, subtitling was only lost for a portion of the programme – on average, just over seven minutes, with more than half being under three minutes;
  - b) technical problems accounted for about 70% of all cases, and operational issues a further 37%;
  - c) 60% of cases occurred when programmes were being subtitled live. 70% of problems with live subtitling resulted from technical causes. This may suggest

that minimising the need for live subtitling would also help to reduce the likelihood of technical failures;

- d) amongst technical errors, a relatively common phenomenon was for subtitling to freeze – something that many viewers will be familiar with. Other common problems included a failure for subtitles to be inserted at the playout stage, and glitches in the software used in the production of live subtitling;
  - e) amongst operational errors, common problems included failures in communication about late changes of plan or the availability of subtitling files, and inaccurate or absent timecodes or subtitling ‘flags’ in programme files, so that subtitling files did not play out; and
  - f) other problems included the loss of network connections, and computer-based equipment (such as channel servers) ‘crashing’.
- 6.21 Ofcom notes that there was no obvious pattern to most technical and operational errors. However, there were clusters of errors which might bear scrutiny by broadcasters, such as subtitles that freeze, subtitling files that were incorrectly coded or flagged, and software glitches. For this reason, Ofcom will ask for a further report covering the period from November 2014 to January 2015. The anonymised results will be summarised as part of the report on live subtitling due in Spring 2015.

## Other matters

- 6.22 The BBC said that its subtitling provider was continuing to roll out new subtitling software across its UK-wide, nations and regions programming. Inevitably, this might result in occasional problems, but in the medium term, should improve quality. The BBC said that it had held a roundtable with groups representing the interests of access service users, and hoped to continue the dialogue.
- 6.23 ITV said that its access service provider had adopted the NER model for internal quality checking, and that ITV and its access services provider had introduced a ‘secret viewer’ panel to regularly monitor the accuracy, latency, presentation and overall experience of subtitles on ITV, and provide constructive feedback each month.

## Annex 1

# External review of measurements



Pablo Romero-Fresco

Inma Pedregosa

30 March 2014

## Live Subtitling: A Qualitative Analysis of the First Round of Measurements

### 1. Methodology

On 16th October 2013, Ofcom decided that broadcasters should be required to measure the following dimensions of quality, on the basis of samples of live subtitling selected by Ofcom:

- a) the average speed of the subtitling;
- b) the average latency of the subtitling (the delay between speech and live subtitling), and the range of latencies; and
- c) the number and type of errors (i.e. minor spelling errors, major omissions or factually misleading subtitles).

Ofcom identified samples of live subtitling in three genres of programming - news, entertainment and chat shows - and asked broadcasters to carry out measurements. Broadcasters were asked to collect data using the NER model, devised by Pablo Romero-Fresco (University of Roehampton) and currently used by regulators, broadcasters and subtitling companies in Australia, Spain, Germany, Switzerland, Italy and France, among other countries.

In order to ensure consistency of measurements, a small team of experts at the University of Roehampton led by Pablo Romero-Fresco and Inma Pedregosa has validated the measurements provided by broadcasters from a third-party standpoint, using the NER model.

The external reviewers have analysed **a total of 66 ten-minute clips** from 66 programmes belonging to three genres (news, entertainment and chat shows) and broadcast on five channels, namely BBC1 News at Six, Strictly Come Dancing and Graham Norton from the BBC, C4 News and Sunday Brunch from Channel 4, Five News from Channel 5, ITV News, X Factor and Alan Titchmarsh from ITV and Sky News at Six and Soccer AM from Sky. In total, the analysis comprises ten hours of live TV material including approximately 103,000 words and almost 15,000 subtitles.

The results of the individual evaluations of every programme are not restricted to a single figure, in this case the accuracy rate. Instead, the NER model provides **a short assessment** of

the quality of the subtitles for every programme, including the accuracy rate and also issues related to the delay of the subtitles, their position, their speed, their flow, the types of errors they contain, the way in which the speakers have been identified and the challenges posed by every programme, among other factors.

As far as the **accuracy rate** is concerned, the present analysis confirms the figure of **98%** as a valid threshold from which subtitles may be considered of acceptable quality for TV broadcast. In our sample, and taking into account also the above-mentioned factors related to overall quality, subtitles with 99%-99.5% accuracy rate ranged from very good to excellent, whereas those with 98.5%-99% ranged from good to very good. Below those rates, the closer to 98% accuracy rate the subtitles were, the more problems they presented. For only two out of the 66 programmes analysed, accuracy rates below 98% were found to be acceptable on the basis of the other elements in the subtitles. In the rest of the programmes, the accuracy rate (and the threshold of 98%) was found to be in line with the overall quality assessment of the subtitles, including delay, position, speed, flow, speaker identification, etc.

A common concern regarding the application of quantitative and qualitative measures to assess the quality of live subtitles is the existence of discrepancies and subjective evaluations, especially when it comes to analysing the loss of information on the subtitles and the impact it may have on the viewers. However, the application of the NER model has proved very consistent and comparable across the internal reviewers from the different subtitling companies (who were only given a few written instructions as to how to apply the model) and the external reviewers from the University of Roehampton. The **average discrepancy** with regard to the accuracy rates of the 66 programmes from five different broadcasters is **0.09%**.

## 2. Results

On the whole, the **quality of the live subtitles** analysed in this report may be regarded as good, with an average accuracy rate of 98.5%, including some very good examples (one fifth of the total sample) and also some subtitles that fall below the required quality threshold (also one fifth of the programmes). This is a considerable improvement from the last analysis conducted by the University of Roehampton at the end of the 2012 London Olympics and it bears witness to the effort made by access service providers and subtitlers to increase the quality of live subtitles on TV, despite the challenging nature of the job. However, there are still some issues that must be addressed in order to avoid instances where live subtitles prevent deaf and hard-of-hearing viewers from having full access to live TV programmes. Some of these issues, which cannot be solved easily given the nature of the programmes and the technology available, will be explained in the following points. Other issues can be tackled and will be covered as part of the section on recommendations at the end of this report.

Before dealing with the results of the review, a mention must be made of the often-discussed issue of **verbatim vs. edited subtitling**. When it comes to live programmes, fully verbatim subtitles are very difficult to produce and often not desirable. Their difficulty is mainly caused by the high speech rates of some of the speakers on TV, often over 180 words per minute (wpm) and reaching peaks of 250wpm and even higher. Only some stenographers are able to transcribe speech literally, and only if there is no overlapping speech, which often features in chat shows, to mention one example. Respeakers will find it even more difficult to provide verbatim subtitles for these programmes, given that they have to dictate not only

what is being said by the speakers but also punctuation marks and other commands. Moreover, fully verbatim subtitles for presenters speaking over 180wpm would not be desirable, since they would cause the viewers to spend between 80% and 90% of their time on the subtitles and only 10%-20% on the images<sup>8</sup>. Thus, for many live programmes, the quality of the subtitles will depend on how good the subtitlers' editing is.

The overall subtitling quality of the programmes included in this sample varies greatly depending on the **genre** analysed, which is mainly due to three factors: the speakers' speech rates, the content and the structure of the programmes.

The **entertainment programmes** analysed here feature the lowest average speech rate (161wpm), they combine live spontaneous dialogue with pre-recorded material and songs (thus providing much-needed breathers for subtitlers) and they lend themselves to a certain degree of editing, since the spontaneous dialogue often features hesitations, false starts and redundant comments. This may explain why despite a fairly high editing rate (an average of 16%), the entertainment programmes in the sample have obtained the best overall accuracy rates of all three genres (98.80%). The sample includes some very good examples with an accuracy rate of over 99% and only a couple that do not meet the minimum requirements in terms of quality.

The **news programmes** in the sample feature a similar average speech rate (163wpm) to entertainment programmes but they normally have more meaningful content, which is particularly challenging to edit without losing important information. Given the importance of these programmes and their content, subtitlers seem to be making a significant and commendable effort to edit less than in entertainment programmes (an average editing rate of 12%), managing to keep the accuracy rate above the required threshold (98.60%). One third of the examples analysed may be considered as very good but a few programmes (13% of the total sample) do not meet the minimum requirements in terms of quality. Some of the news programmes analysed combine live and semi-live subtitles for the parts where the scripts are available to the subtitler before transmission. The latter sections are thus shown with synchronised, verbatim subtitles, which is a welcome addition. Yet, there is still some work to be done on the transition between these two parts, which often involves a considerable delay in the live part or an excessive speed when live subtitles give way to pre-prepared subtitles so that they can catch up with speech. It may be worth considering combining two different display modes, as is the case in Italy, making a distinction between live subtitles (scrolling) and pre-prepared subtitles (blocks) and introducing the pre-prepared blocks gradually with some delay until they are completely synchronous. This would ensure that they do not exceed the required reading speed for the viewers. Other problems that have been identified in the news programmes analysed are the lack of speaker identification in some cases, subtitles that obscure the speakers' mouths (thus making it impossible for viewers to lip read) and subtitles containing warnings for disturbing graphic material that are displayed after the graphic material has been shown. Given the unavoidable delay in live subtitling, it may be necessary to include these warnings as part of the visual text on the programme.

**Chat shows** are often the most difficult genre to tackle as far as live subtitling is concerned, not only because of the speakers' fast speech rates (an average of 180wpm in this sample, with peaks of 250wpm) but also because of the presence of overlapping speech that is essential to understand the humour in the programme. The subtitlers have thus been forced to

---

<sup>8</sup> Romero-Fresco, Pablo (2011) *Subtitling through Speech Recognition: Respeaking*, Manchester: St. Jerome.

edit an average of 22% of the original content and, on the whole, have just managed to reach the accuracy threshold (98.15%). Only one fifth of the programmes contain very good subtitles but more than one third do not meet the minimum requirements in terms of quality. In some programmes, the editing could have been more effective if subtitlers had managed to leave out unimportant comments, focusing instead on those remarks that carry the comic weight of the programme. For the many occasions on which this is impossible, it may be advisable to use a symbol, for example (...), as is the case in Canada, so that viewers know that some information is missing in the subtitles. Indeed, one of the most important factors regarding viewers' comprehension is whether or not they are aware that an error (or an omission) has occurred, which could be made clear with the use of this symbol. Finally, presenters in chat shows often refer to pictures and photographs shown on the screen. Given that the subtitles are likely to be displayed with delay, it may be a good idea for subtitlers to replace deictic elements such as "this", "that", "here" and "there" with more explicit words to help viewers understand what the presenters' comments are referring to.

The overall **latency** of the programmes analysed is 5.6 seconds. As shown in the sample analysed here, the delay of live subtitles depends largely on the availability and use of pre-prepared scripts, the live subtitling technique used, the genre of the programmes and the occurrence of specific technical issues during transmission. When scripts are available and pre-prepared subtitles are combined with live subtitles, the average delay has been reduced to 3-4 seconds and on rare occasions even to 2-3 seconds. As was expected, subtitles made by stenographers have less delay (an average of 3-5 seconds) than those made by respeakers, which are closer to 6 and often 7 seconds. On the whole, programmes with many speakers, fast speech rates and overlapping speech are likely to cause delays of over 7 seconds, which can have a significant impact on viewers' comprehension. Finally, regardless of the average delay of the subtitles and possibly due to specific technical problems, several programmes in the sample analysed feature peaks of 10-second delays and sometimes even more, which should be avoided whenever possible.

In the sample analysed, 63% of the **errors** included in the subtitles were editing errors (those caused by incorrect omissions made by the subtitlers) and 37% were recognition errors, those caused by the interaction between the subtitler and the steno machine or the speech recognition software. Once again, these percentages vary depending on the genres. In chat shows, given the fast speech rates of the speakers, 70% of the errors are caused by incorrect editions and 30% by misrecognitions. Entertainment programmes, which feature slower speech rates, contain 65% edition errors and 35% recognition errors. Finally, news programmes feature 55% edition errors and 45% recognition errors, the latter perhaps due to the effort made by subtitlers to type/respeak fast in order to keep up with the audio without editing too much, thus causing more recognition errors. As far as live corrections are concerned, in general subtitlers are advised to disregard minor errors (as correcting them only adds delay to subsequent subtitles in cases in which comprehension has not been affected in the first place) and to correct serious errors (those that introduce misleading information).

### 3. Specific issues

This section includes the different types of issues found in the sample analysed.

- Editing Choices:

In many programmes, (correct) editions were often necessary to “keep up with the merciless pace” of dialogue, as one reviewer put it. Some subtitlers seemed to do this consciously and effectively, whereas in other cases there were poor editing choices, normally due to excessive edition or unnecessary rewording. In programmes featuring recipes, it is important to convey all processes so that viewers can understand them and replicate them at home. Similarly, in other themed programmes (dancing, singing), it would be advisable to keep all comments related to the contestants' abilities whenever possible. In football programmes with commentaries over pre-recorded highlights, these should have been subtitled as a football match normally is, i.e. subtitling only those comments that add to what is being seen in the images. Subtitlers have the extra task to match the subtitles with images and text on screen and a few times, the editing of the speech when there was text on the screen was particularly impressive. However, deictic elements such as "this" and "that" should have been avoided and explicitation included in sentences such as "I was doing that" (meaning *I was looking at my phone*) and "You do this" (meaning *You tuck your chin in*).

- Omissions:

Good and excellent subtitles showed clever omission of some unnecessary words and sentences. In contrast, a news programme featured consistent problems of omission of information when the national coverage split to regional news. Several programmes avoided foreign names and proper nouns in the subtitles, presumably to avoid errors in cases where these terms had not been anticipated or prepared beforehand.

- Problems with recognition:

In some cases, misrecognitions could have been avoided with better preparation, examples include “Lady GaGa” (*Lady Gaga*), “Present Cesc lay ya” (*Princess Leia*) and “Madeline” (*Madeleine*). Furthermore, in specific programmes or sections (e.g. football, recipes), more attention should have been paid to avoid content-specific errors such as “sauce” (*source*) or “lepl on” (*lemon*). The subtitles of some programmes contained commands such as “capital letter”, “full stop”, “question mark”, etc. and there was one whole section of a programme that came out entirely in upper case.

- Spelling and punctuation:

Some channels seem to be more meticulous than others in the use of orthotypographic conventions such as commas, quotation marks and capitalisation. The use of contractions, which may be useful to convey the manner of speaking, was not widely spread. This may be explained by the fact that speech recognition software works better with full words ("we are") than with contractions ("we're"), so some subtitlers may have chosen to avoid the latter. Some programmes had a problem with the representation of apostrophes (the subtitles contained a space instead of the apostrophe). Although the meaning was clear, it was distracting.

- Corrections:

Although a few programmes featured timely and successful corrections, generally there was little attempt to correct errors and in some cases they were displayed on the screen quite late. Serious recognition errors, which introduce misleading information, should be corrected (examples include "they need a man" instead of *they need a mum*, "be given to ayatollahs" instead of *be given to our toddlers*, etc.).

- Latency:

As explained above, the overall latency of the programmes analysed was 5.6 seconds and many programmes featured long stretches with short, steady latency. However, most

programmes (regardless of the channel and the genre) also contained higher peaks. In some cases, these peaks reached delays of 9, 15, 17, 18 and even 24 seconds. In the majority of cases, these long delays caused omission and misreporting of information and facts, rewording of sentences in a way they became misleading and loss of references to the images (photos or videos) appearing on the screen at the time of the original speech.

- Cued subtitles:

Pre-prepared subtitles contributed greatly to lower the subtitle latency. During cued sections there was almost no latency and subtitles were almost verbatim. However, on three different channels, the subtitles were displayed before the images and on a couple of occasions, two subtitles were introduced which did not correspond with the speech, probably because they had been based on an earlier version of the script.

- Technical issues and absence of subtitles:

The technical issues reported involved an issue with the speech recognition software that led the subtitler to revert to typing and a non-specified technical problem that triggered repeated text. Freezes and absence of subtitles were identified in several programmes on different channels. There were stretches of 7, 13 and 40 seconds (plus two other long stretches, of non-specified duration) without subtitles and freezes of 22 seconds. There were also instances of subtitles appearing and/or disappearing suddenly and, as explained above, at times subtitles were displayed very quickly in order to catch up with the script, increasing the reading speed to unattainable levels. The sample analysed also contained erroneous repetition of full subtitles, clauses and words.

- Captions and images obscured:

There were many instances of speakers' mouths being obscured by subtitles. This happened across all channels, genres and programmes and it is especially problematic for viewers who lip read. Most of the examples found in the analysis could have been avoided by raising or lowering the subtitles, as no other text was shown on the screen. There were also some instances of on-screen information being obscured by the subtitles.

- Speaker identification:

The use of colours to identify the speakers was normally consistent and effective. However, there were also programmes with no speaker identification (which has a significant impact on the viewers' comprehension), news programmes where the same colour (white) was used for all reporters and cases in which the same person was subtitled with two colours or two people were subtitled with the same colour. The use of colour green in some programmes to identify one of the main speakers is problematic, as this colour is less legible and more tiring to read than white, yellow or cyan.

- Use of labels:

There was good use of labels (SCREAMS or Translation:) in some programmes but the majority of them did not make use of this feature, which contributes to comprehension.

- Censorship:

There was an issue with the word "paedophile" (uttered after the watershed in an entertainment show), which was said twice but was not included in the subtitles. Likewise, an expletive ("bloody") before the watershed was carefully omitted.

#### **4. Recommendations**

This section includes several recommendations that may help to tackle some of the issues identified in the analysis:

- The use of two different display modes in programmes containing a combination of live subtitles (displayed in scrolling mode) and pre-prepared subtitles (which could be displayed in blocks). This practice has been implemented successfully in Italy for some years now and it enables viewers to spend more time on the images and less time on the subtitles when pre-recorded subtitles in blocks are displayed. More attention needs to be paid to the transition between the sections with semi-live and live subtitles, which at the moment is causing a considerable delay in the live part or an excessive speed when live subtitles give way to pre-prepared subtitles. It may be useful to introduce the pre-prepared blocks gradually with some delay until they are completely synchronous. This would ensure that they do not exceed the required reading speed for the viewers.
- The use of a symbol, for example (...), to indicate omission of information in instances where the viewers need to know that important information has been left out of the subtitles.
- More attention, whenever possible, to the content on the screen, which is as important as challenging and which is often the sign of a good and skilled subtitler. This attention is essential to avoid redundancy in the subtitles and especially to avoid obscuring speakers' lips and important textual information. Also important here is to avoid deictic references (“this”, “that”, “here” and “there”) when speakers refer to pictures and photographs on the screen if the subtitles are going to be displayed with the usual delay. More explicit terms may have to be used to refer to those visual elements.
- The use of additions, which can sometimes be an effective tool to help comprehension: e.g. “It is so simple to do, and (*it adds*) loads and loads of flavour“.
- The correction of serious errors (misleading information) but not of minor errors, since the latter would only add delay to subsequent subtitles in cases in which comprehension has not been affected in the first place. Standard recognition errors (i.e. nonsense) may be corrected when dealing with key information and/or if there is time to do so without incurring in too much delay. As for standard edition errors (i.e. full factual omissions), the use of the symbol (...) may be helpful in instances where the viewers need to know that important information has been left out of the subtitles.
- The use, whenever possible, of short, full sentences instead of coordinated sentences linked by “and” and “but”. These short sentences are easier to follow by the viewers and often help the subtitler to catch up with the original audio.
- The introduction of textual content warnings as part of the programme instead of the subtitles. At the moment, subtitles containing warnings for disturbing content are often displayed after the images have been shown, given the unavoidable delay in live subtitling. This could also apply to warnings for flash photography, strobe lighting, etc.

## Annex 2

# Application of the NER model to measuring accuracy

A2.1 The following table illustrate the application of the NER model to measure the accuracy of live subtitling in three excerpts from news programmes. The transcript of the original speech is in column one, and the subtitling is in column two. The third column lists all the errors identified in the segment, which are used then to calculate the NER score – i.e. the ‘accuracy rate’, plus all the correct editions done by respeakers.

A2.2 It should be noted that these excerpts are here reported only for illustrative purposes. They do not represent the entirety of the sample, nor do the accuracy rates shown here reflect the overall accuracy of subtitling in the programmes from which they were taken.

A2.3 Key for the abbreviations<sup>9</sup>:

- MinE – *Minor Edition Error*
- StE – *Standard Edition Error*
- SE – *Serious Edition Error*
- MinR – *Minor Recognition Error*
- StR – *Standard Recognition Error*
- SR – *Serious Recognition Error*
- CE – *Correct Edition*

ORIGINAL SPEECH	SUBTITLES	ERRORS	ACCURACY RATE (NER)
I'm joined now from Washington by Dean Clancy, Vice-President of public policy for the grassroots organisation Freedom Works which supported the Conservative tea party movement and opposes Obamacare.	We're joined now from Washington by Dean Clancy, vice president of Policy for the grassroots organisation FreedomWorks, which supported the Conservative Tea Party movement and opposes Obamacare. -- vice president of public policy. <sup>1</sup>	1. Vice president CE	N = Number of Words = 254 E = Edition Errors = 0.25+0.25+0.25+0.5 = 1.25 R = Recognition Errors = 0.25+0.25+1 = 1.5 NER = (N-E-R)/N = (254-1.25-1.5)/254 = 251.25/254 = <b>98.9%</b>
He was a senior	He was a senior		

<sup>9</sup> For a detailed explanation of the different categories of subtitling error, refer to: <sup>9</sup> Pablo Romero-Fresco and Juan Martínez, *Accuracy Rate in Live subtitling – the NER Model* (2011), Roehampton University, UK (<http://roehampton.openrepository.com/roehampton/bitstream/10142/141892/1/NER-English.pdf>)

<p>adviser on Capitol Hill and sat in on negotiations during the last US shutdown. But, Mr Clancy, let me congratulate you first for America joining the rest of the civilised world, you've at last got a comprehensive healthcare system and that hasn't been affected by the shutdown today.</p> <p>Many congratulations, are you thrilled? No, not at all, but thank you. Obamacare is a bad thing. It's unaffordable. It's driving up people's healthcare costs and it's putting politicians and bureaucrats in charge of our healthcare.</p> <p>So it's nothing to celebrate. You know, we think that markets can work in healthcare and that any problems in healthcare are the result of government intervention rather than markets. So actually this is a pretty sad day, but the real beginning of Obamacare is not today it's –</p> <p>Let me pause you there. Sad day. You say healthcare is unaffordable. Let me tell you that the rest of the world finds what America is doing, you legislators in</p>	<p>advisor on Capitol Hill and sat in on negotiations during the last US shutdown. Let me congratulate you first for America joining the rest of the civilised world. You have at last got a comprehensive health care system, and that has not been affected by the shut down today. --<sup>2</sup></p> <p>Many congratulations, are you thrilled? Not at all, <sup>3</sup> thank you. Obamacare is a bad thing. It is unaffordable, driving up health care costs and putting politicians and bureaucrats in charge of health care.</p> <p>It is nothing to celebrate. We think that markets work<sup>4</sup> in health care and any problems in health care are the result <sup>5</sup> government intervention rather than market. So this is a pretty sad day <sup>7</sup>.</p> <p>Let me pause you. Sad day. You say health care is unaffordable. The rest of the world finds what America is doing, you legislators in</p>	<p>2. -- many congratulations, MinR (0.25) [<i>double dash.</i>]</p> <p>3. Not at all but thank you/Not at all, thank you – MinE (0.25)</p> <p>4. Markets can work/markets work – MinE (0.25)</p> <p>5. Of. MinE (0.25)</p> <p>6. rather than markets/rather than market. – MinR (0.25)</p> <p>7. , but the real beginning of Obamacare is not today it's [...] StE (0.5) [<i>This idea has been omitted</i>]</p>	
--	--	---	--

<p>particular, or the ones you support as doing, as unaffordable for us too.</p>	<p>particular or the ones who supported<sup>8</sup>, as unaffordable for us, too.</p>	<p>8. The ones you support/the ones who supported – SR (1) [<i>wrong info</i>]</p>	
<p>We've changed her voice to protect her identity. It's like being in a silent world. You don't tell anybody because you don't want people to know. She says it's not as simple as just asking a victim if they're being abused. You would probably get women who are still frightened and still not have the confidence to actually say something is going on.</p> <p>So nurses at this hospital in Rochdale are being trained to ask the right questions and spot the warning signs. They'll work closely with victim support and police. The first signs that they may show of having been abused will be when they turn up in the A&amp;E department. So the level of training there, the intervention there makes an awful lot of sense. But what if staff jump to the wrong conclusions and partners are wrongly accused?</p>	<p><sup>1</sup> Luke<sup>2</sup> it is like being in a silent world, you don't tell anyone because you don't want anyone to know. She says it is not as simple<sup>3</sup>. You would still get women who are frightened and he would not have the confidence<sup>4,5</sup>.</p> <p>Nurses<sup>6</sup> are being trained to ask the right questions and spot the warning signs,<sup>7</sup> they will work closely with victim Support<sup>8</sup> and the police. The first signs of abuse<sup>9</sup> may turn up in the A&amp;E department. <sup>10</sup> That early intervention makes an awful lot of sense. But what his<sup>11</sup> staff jump to the wrong conclusions and partners are accused<sup>12</sup>?</p>	<p>1. We've changed her voice to protect her identity. StE (0.5) [<i>this is a full independent idea unit</i>]</p> <p>2. Luke MinR (0.25) [<i>insertion</i>]</p> <p>3. She says it is not as simple. (She says it is not as simple as asking a victim if they are being abused.) StE (0.5)</p> <p>4. ..and he would not have the confidence (..and who would not have the confidence) StR (0.5)</p> <p>5. ..to actually say something is going on. StE (0.5)</p> <p>6. at this hospital in Rochdale StE (0.5)</p> <p>7. spot the warning signs, (.) MinE (0.25) [<i>inappropriate punctuation</i>]</p> <p>8. victim Support MinR (0.25) [<i>capitalisation</i>]</p> <p>9. The first signs of abuse may turn up (The first signs that they may show of having been abused will be when they turn up) SE (1)</p> <p>10. The level of training there StE (0.5)</p> <p>11. But what his staff (But what if staff) StR (0.5)</p>	<p>N = Number of Words = 115</p> <p>E = Edition Errors = 0.5+0.5+0.5+0.5+0.25+1+0.5+1 = 4.75</p> <p>R = Recognition Errors = 0.25+0.5+0.25+0.5 = 1.5</p> <p>NER = (N-E-R)/N = (115-4.75-1.5)/115 = <b>94.56%</b></p>

		<p>12. ..and partners are accused? (..and partners are wrongly accused?) SE (1) [<i>vital word edited out</i>].</p>	
<p>Today the inquest into the death of Mark Duggan heard from the marksman who fired the fatal shot. Members of Mark Duggan's family were in court, to see for the first time the officer who killed him. Lawyers and the jury could also see him but to protect his identity, the press and the public sat in a separate courtroom where they could hear the officer's testimony but couldn't see him.</p> <p>The marksman known as a V53, described how he was part of a specialist firearms unit, which was following Mark Duggan, who was in a minicab. The police intelligence was that he had just collected a gun. V 53 was in the third police car, he said that in Ferry Lane in Tottenham, the lead officer in the first car decided they should carry out what is called a hard stop on Mark Duggan's cab.</p> <p>He described how surrounded by police, Mark Duggan jumped out of the minicab and V53 believed he was</p>	<p>Today, the inquest into the death of Mark Duggan heard from the marksman who fired the fatal shot. Members of Mark Duggan's family were in court to see for the first time the officer who killed him. Lawyers and the jury could also see him, but to protect his identity, the press and public sat in a separate court room, where they could hear the officer's test me<sup>1</sup> but could not see him.</p> <p>The marksman, known as V53, described how he was part of a specialist firearms unit which was following Mark Duggan, who was in a mini cab unit<sup>2</sup>. The police intelligence was that he had just collected a gun. V53 was in the third police car. He said that in this road<sup>3</sup> in Tottenham, the lead officer<sup>4</sup> decided they should carry out what is called a hard stop on Mark Duggan's<sup>5</sup>.</p> <p>He described how, surrounded by police, Mark Duggan jumped out of the minicab, and V53</p>	<p>1. Officer's test me: StR (0.5)</p> <p>2. Unit: MinE (0.25) [extra word inserted]</p> <p>3. Ferry Lane: MinE (0.25) [dependent idea unit]</p> <p>4. In the first car MinE (0.25)</p> <p>5. Cab MinE (0.25)</p>	<p>N = Number of Words = 316</p> <p>E = Edition Errors = 0.25+0.25+0.25+0.25+0.5 = 1.5</p> <p>R = Recognition Errors = 0.25</p> <p>NER = (N-E-R)/N = (316-1.5-0.25)/316 = 314.25/316 = <b>99.45%</b></p>

<p>looking to escape. He told the jury Mr Duggan turned to face him and he was sure he had a gun. He Said, "The world has just stopped in my head. It's like a freeze frame moment. The only thing I was focussing on is the gun. He's raised the weapon, moved it a couple of inches away from his body... I know have an honestly held belief that he is going to shoot me."</p> <p>It was then that V53 opened fire. No gun was found near to Mark Duggan's body but a weapon was discovered on the other side of some railings on grass. It was covered in a black sock. During cross-examination by Leslie Thomas, one of the barristers representing Mark Duggan's family, it was put to V53 that the reason no weapon was found close to the body was because Mark Duggan did not actually have a gun. The officer replied, "That was complete rubbish."</p>	<p>believed he was looking to escape. He told the jury that Mark Duggan turned to face him, and he was sure he had a gun. He said [information on screen]<sup>6</sup>...</p> <p>It was<sup>8</sup> then that V53 opened fire. No gun was found near to Mark Duggan's body, at a weapon<sup>7</sup> was discovered on the other side of some railings on grass, covered in a black sock. During cross-examination by Leslie Thomas, one of the barristers representing Mark Duggan's family, it was put to V53 that the reason no weapon was found close to the body was because Mark Duggan did not actually have a gun. The officer replied, that is complete rubbish.</p>	<p>6. The world has just stopped...the gun: StE (0.5) [although the sentence appears on screen, there are still scrolling subtitles, and it is not possible to read both at once]</p> <p>7. At a weapon: MinR (0.25)</p> <p>8. It was: CE</p>	
---	---	---	--